Systematization of Knowledge on Hardening Kubernetes through Native Configurations

LABCIB

Master in Informatics Engineering - 2024/2025

Porto, December 30, 2024

1190830 - Luís Correia
1190974 - Pedro Lemos
1191526 - Pedro Moreira

Version 2, 2024-10-17

Revision History

Revision	Date	Author(s)	Description
1	2024-11-15	Pedro Moreira	Initial version and structure
2	2024-10-17	...	Extended description

Contents

List of Figures **vi**

1 Introduction **1**

2 Objectives and Scope **3**

3 Methodology **5**

4 Background **7**

4.1 Kubernetes Components 7

4.2 What is etcd? 8

4.3 What are Kubernetes Secrets? 8

4.3.1 Key Features of Secrets 9

5 Kubernetes Native Configurations for Hardening **11**

5.1 Data Encryption 11

5.2 API Server Configurations 13

5.2.1 Authentication 14

5.2.2 Admission Controllers 16

5.2.3 API Server Security Flags 20

5.3 Network Policies 24

5.3.1 Policy Types 25

5.3.2 Component Analysis 26

5.3.3 Best Practices 27

5.4 RBAC and Authentication 27

5.4.1 Role and ClusterRole 28

5.4.2 RoleBinding and ClusterRoleBinding 28

5.4.3 Implementation of RBAC 29

6 Complementary Tooling **31**

6.1 Falco 31

6.2 OPA (Open Policy Agent) 32
6.3 Trivy 33
6.4 Vault (by HashiCorp) 33
6.5 Kube-bench 36

7 Analysis and Best Practices **39**

8 Conclusions **41**

List of Figures

5.1 Access Control Overview [**the_linux_foundation_controlling_2023**] 16
5.2 Enabled admission controllers in testing instance 17
5.3 Configuration of API server for audit logs 22
5.4 Audit logs of a namespace creation 23
5.5 Audit logs of a pod creation 24

6.1 Trivy Summary Report [**trivy_doc_kubernetes_nodate**] 34
6.2 Vault Flow [**vault_doc_introduction_nodate**] 35
6.3 kube-bench output example [**noauthor_kube_bench_nodate**] 36
6.4 kube-bench remediation output [**benhassan_depth_2024**] 37

1 Introduction

The topic of configuring and managing Kubernetes clusters has a significant importance in the context of modern technological infrastructures. Kubernetes, recognized as one of the leading tools for container orchestration, offers a robust platform for deploying scalable and resilient microservices. This study delves into the configurations of Kubernetes components, performance optimizations, and the application of security best practices, particularly in secrets management and data encryption.

The increasing reliance on Kubernetes in corporate environments underscores the need for secure and efficient solutions. By providing a comprehensive theoretical framework alongside practical implementation strategies, this work seeks to address key challenges and propose effective methodologies for leveraging Kubernetes to meet contemporary organizational demands.

2 Objectives and Scope

[Description, for what under individual responsibility, that inform the goal, the questions related to functional correctness, metrics used and their values, with partial response to questions, and goal achievement analysis. Explicitly mention the tool report(s).]

3 Methodology

[Description, for what under individual responsibility, that inform the goal, the questions related to maintainability, metrics used and their values, with partial response to questions, and goal achievement analysis. Explicitly mention the tool report(s).]

4 Background

Kubernetes is a container orchestration platform that automates the deployment, scaling, and management of containerized applications. Originally developed by Google and now maintained by the Cloud Native Computing Foundation (CNCF), Kubernetes enables IT teams to deploy and manage applications at scale efficiently using containers.

4.1 Kubernetes Components

Kubernetes architecture is based on a distributed system that consists of two primary types of nodes: master nodes and worker nodes. Each type plays a specific role in ensuring the cluster operates effectively [**the_linux_foundation_kubernetes_2024**].

Master Nodes

Master nodes are the control center of a Kubernetes cluster. They manage the cluster's state, handle scheduling, and coordinate operations. The key components running on master nodes are:

- **API Server** (`kube-apiserver`): Exposes the Kubernetes API and serves as the entry point for all cluster interactions.
- **Controller Manager** (`kube-controller-manager`): Manages controllers that monitor the clusters state and make adjustments to achieve the desired state.
- **Scheduler** (`kube-scheduler`): Component that watches for newly Pods without assigned nodes, and selects one for them to run on. Some factors are taken into the decision, like hardware/software/policy constraints, data location, deadlines, among others.
- **etcd**: Stores all configuration data and the state of the cluster in a consistent and reliable way.

- **Cloud Controller Manager** (`cloud-controller-manager`): This component runs cloud-provider-specific controllers that manage resources outside the Kubernetes cluster. It is only active when the cluster is hosted on a cloud platform.

Worker Nodes

The main responsibilities of a worker node is to process data stored within the cluster and managing networking to ensure seamless communication both between applications within the cluster and with external systems. The control plane assigns tasks to the worker node to carry out these functions. The key components are:

- **Kubelet**: An agent that ensures containers are running and in healthy state according to specifications.
- **Kube Proxy**: Manages network routing and load balancing at the network level.
- **Container Runtime**: The software responsible for running containers.

4.2 What is etcd?

etcd is a distributed key-value database essential for storing the cluster's state and configuration. It ensures data consistency across all nodes using the Raft consensus algorithm, which keeps all nodes in sync even during failures, providing a reliable foundation for Kubernetes operations [**etcd_faq_2024**].

A leader is elected among the etcd nodes to manage all write operations, while follower nodes stay synchronized with the leader. If the leader fails, the Raft algorithm ensures a new one is elected quickly, maintaining high availability and preventing service disruption [**etcd_faq_2024**].

Additionally, etcd persistently stores the cluster's state, allowing Kubernetes to retrieve and update this information in real-time. It holds metadata and operational data for Kubernetes objects like pods, deployments, and services, enabling the Kubernetes API server to query etcd and consistently enforce the cluster's desired state [**etcd_faq_2024**].

4.3 What are Kubernetes Secrets?

In Kubernetes, a **Secret** is an object designed to securely store sensitive data, such as passwords, API tokens, SSH keys, or other confidential information. Secrets help prevent the exposure of sensitive data by avoiding hardcoding such information directly in application code or configuration files [**the_linux_foundation_secrets_2024**].

Kubernetes Secrets were introduced to provide a more secure and flexible approach to managing sensitive data. Secrets are stored within the Kubernetes cluster and can be accessed by containers running in the cluster without exposing the data directly in code or static files. While they are base64 encoded by default, additional measures, such as encryption, can be applied to protect them [**the_linux_foundation_secrets_2024**].

4.3.1 Key Features of Secrets

- **Storage**: Secrets are base64 encoded before being stored in `etcd`, but not encrypted by default. However, Kubernetes can be configured to encrypt Secrets at rest using tools like Key Management Service (KMS).
- **Usage**: Secrets can be mounted as volumes in containers or passed as environment variables. This method allows for secure and flexible handling of credentials without exposing them in plain text.
- **Access Control**: Kubernetes uses Role-Based Access Control (RBAC) to define which users or applications can access specific Secrets. This ensures that only authorized entities can retrieve sensitive information.
- **Decoupling Configuration**: Sensitive data is decoupled from application code and deployment configurations. This simplifies updates to credentials without requiring changes to the underlying codebase.

Kubernetes supports several types of Secrets to cater to different use cases. The most common type is the **Opaque Secret**, which allows the storage of arbitrary key-value pairs, making it suitable for storing passwords, tokens, or generic credentials. For handling Docker credentials, Kubernetes offers **Dockercfg** Secrets, which store authentication details for accessing private container registries. Finally, there are **TLS Secrets**, specifically designed to store TLS certificates and their corresponding private keys, enabling secure communication between services within the cluster [**the_linux_foundation_secrets_2024**].

5 Kubernetes Native Configurations for Hardening

Having introduced what Kubernetes is, we are now able to delve into the main purpose of this report, which is to systematize what built-in features and configurations of this technology we can use to improve a cluster's security posture and resistance to cyber-attacks. This chapter will be split into various sections, each concerning an area of relevance when it comes to Kubernetes Hardening. In each section, a context regarding the security improvements that are available will be presented, and consequently the measures through which one can implement these improvements will also be showcased.

5.1 Data Encryption

As we've seen in the context section of this report, Kubernetes stores Secrets and their data on the etcd backend database without any form of encryption [**the_linux_foundation_secrets_2024**] - despite encoding them with Base64, which does not constitute encryption as the plain-text contents are directly obtainable. If an attacker is able to access either the API server or the etcd database directly, he will be able to obtain the plain-text contents of the secrets that are managed by that cluster. One can easily picture the compromise that this can cause, as secrets usually contain business-critical credentials that are used by services running in Kubernetes. For instance, an application Pod might need to access a database, and thus read the connection string and database user credentials from a secret. If an attacker could obtain the content of this secret, he too would at least have the credentials to access the database.

Fortunately, Kubernetes provides a series of configurations that can be applied to encrypt Secrets - and other resources - at rest. This is done through the configuration of some resources on the cluster's API server. Notably, Kubernetes defines the `EncryptionConfiguration` Custom Defined Resource (CRD), which allows practitioners to customize what resources shall be encrypted and through what mechanism (also known as provider) [**the_linux_foundation_encrypting_2024**].

The flow of actions is as follows: a user or application creates a Secret by sending an API call to the Kubernetes API; the API checks its `EncryptionConfiguration` file to evaluate what encryption provider to use and which key to use for that provider; if a provider is specified, as well as the key to use, the API server uses both to encrypt the provided data; then, this encrypted data is stored in etcd. For the decryption, the API retrieves the secret from etcd and uses the aforementioned `EncryptionConfiguration` to decrypt the contents.

In 5.1, we present a simple example of an `EncryptionConfiguration`. What this does is instruct the Kubernetes API that all Secrets must be encrypted using the `aescbc` provider, using the encryption key specified in `keys.secret` field.

Listing 5.1: EncryptionConfig example

```
apiVersion: apiserver.config.k8s.io/v1
kind: EncryptionConfiguration
resources:
- secrets
providers:
- aescbc:
  keys:
    - name: key1
      secret: <BASE 64 ENCODED SECRET>
- identity: {}
```

There are several providers which we can use when encrypting secrets, which are specified in the Kubernetes documentation itself [1]. Some facts are worth noting:

- The `identity` provider does not encrypt data. It's the default used by Kubernetes, as we've covered before. Despite this, it has practical utilities, such as serving as one of the encryption providers for instance when migrating clear-text secrets to encrypted ones. So, despite not encrypting data, it would be improper to discard its existence;

- The `aescbc`, `aesgcm` and `secretbox` providers require the encryption key to be specified in the encryption configuration file. Using them protects a cluster in the event of a compromise, but if the hosts in which the Kubernetes API is running are themselves compromised, this does not provide additional security, as the attacker will be able to obtain the encryption key from the local filesystem [**gkatziouras_kubernetes_2024**]. Nonetheless, it's an improvement from the default configuration.

[1] This comparison can be seen in this documentation link

Besides the providers mentioned above, there exists the `kms` provider, which is currently in version 2 - and version 1 is deprecated. This provider offers a more complex - but more secure - encryption solution. It's purpose is to solve the issue of having the encryption key defined in a local file, as we've covered previously. To address this matter, the kms provider calls upon two new components [**gkatziouras_kubernetes_2024**]:

- An external Key Management Service (KMS);
- A Kubernetes KMS provider plugin, which is responsible for connecting with the external KMS for the encryption and decryption processes.

This mechanism adopts an "envelope encryption scheme using a two-key approach" [**gkatziouras_kubernetes_2024**], with those keys being: the Data Encryption Key (DEK), which is used to encrypt the resources referenced in the `EncryptionConfig`; and the Key Encryption Key (KEK), which is used to encrypt the DEK and is hosted remotely on the KMS. This way, having the encryption key on a separate system, the main downside of the previous solutions is addressed, since a filesystem compromise of the Kubernetes master nodes wouldn't directly lead to a compromise of the secrets stored in the cluster.

However, we must note that the usage of the `kms` provider requires the usage of third-party tooling, not native to Kubernetes - since it requires the usage of an external KMS.

5.2 API Server Configurations

The API server is an essential component of Kubernetes, as we've explained in 4.1. It is responsible for handling all administrative operations and communications within the cluster, and thus acts as the primary interface for practitioners to interact with the Kubernetes environment. Given its critical role, the API Server is a frequent target for attackers aiming to exploit vulnerabilities, gain unauthorized access, or disrupt cluster operations.

Securing the API server is, thus, a crucial task that cluster administrators must undergo to ensure the integrity, confidentiality and availability of this environment. Misconfigurations or weak security settings can expose the cluster to numerous risks, such as unauthorized access to data, or disruption of operations.

Therefore, to assist practitioners in undergoing this process, we will look into relevant API server configurations which will improve its security posture.

Before proceeding, it's important to note that the API server can be interacted with through many means, such as the `kubectl` utility, as well as REST requests. This highlights the relevance of appropriate firewall practices. Access to the Kubernetes API must be properly limited to business-related needs. A good approach for this is to allow only an organizational VPN's IP or set of IP addresses to make requests to the API, thus limiting the attack vectors. However, were this firewall or VPN to be compromised, further security measures are necessary.

5.2.1 Authentication

When a user attempts to interact with the API server, he must first undergo a process of authentication. This verifies the identity of a user or service before allowing them to communicate with the API server. Kubernetes supports multiple authentication methods which can be employed [**the_linux_foundation_authenticating_2024**].

X.509 Client Certificates

Kubernetes is able to use certificates for authentication purposes. For this effect, the API server must be started with the `-client-ca-file` flag which should point towards one or more valid Certificate Authority (CA) files, which will be used to validate the certificates used by clients when issuing requests to the API [**the_linux_foundation_authenticating_2024**]. These certificates can be generated manually using tools such as `easyrsa` and `openssl`.

With this method "any user that presents a valid certificate signed by the cluster's certificate authority (CA) is considered authenticated" [**the_linux_foundation_authenticating_2024**]. A more proper way to use this is to have users generate their own private key, as well as a Certificate Signing Request (CSR) which specifies necessary attributes that a user certificate must have. Notably, the subject of the certificate will be used by Kubernetes to identify the username of a client, and thus using it for authorization purposes. Then, this CSR is validated and signed by the administrator, using the CA of the cluster, and the certificate is distributed to the client.

This ensures that each user has an account, and further ensures traceability and accountability of any user's actions. However, the manual distribution of certificates might be a hindrance on organizations which are larger and more complex. Nonetheless, it's a native setting of Kubernetes that ensures the hardening of the authentication process.

Service Account Token

These are signed JSON Web Tokens (JWT) which Kubernetes generates so that entities can communicate with the API server [**the_linux_foundation_authenticating_2024**]. These tokens are associated with ServiceAccount objects, which by default represent the identity of an application in Kubernetes [**martin_hacking_2021**] - their function will be expanded upon in the RBAC section. It's unusual that these tokens are used for the authentication of users, as they are more frequently used by Pods which require to communicate with the Kubernetes API, or by processes that run outside of the cluster but frequently need to talk with its API.

OpenID Connect (OIDC)

This authentication mechanism relies on a third-party identity provider, such as Google or Keycloak, for the generation of a JWT which the user can then present to the API server in order to issue requests [**the_linux_foundation_authenticating_2024**]. It's still relevant in the scope of this Systematization of Knowledge, as even though it resorts to an external mechanism, it uses a configuration that's available in Kubernetes itself.

It's worth noting that, unlike workloads - which are represented in Kubernetes via Service Accounts - human users are not represented by Kubernetes objects [**the_linux_foundation_authenticat martin_hacking_2021**]. Therefore, integrating with an external source-of-truth for users such as the ones mentioned above is optimal for handling user interaction with the API Server. The applicability of this can be easier when using a cloud instance, as practitioners can check what options their cloud provider offers [**martin_hacking_2021**], but technologies like Keycloak or Dex enable organizations to deploy an identity management system on their own infrastructure.

Webhook Token Authentication

Users are able to specify authentication tokens on their requests to the API, which must contain the value of this token on the Authorization header of the request. What this authentication method relates to is the assessment of the validity of the provided token.

It works by using a remote webhook service when determining user privileges [**the_linux_foundation_webh** This external service responds to REST requests and allows or denies access to the requested resources. It's essentially a way to delegate the authorization process to a third-party, thus allowing administrators to inject their custom logic as part of an authorization policy [**eknert_kubernetes_2022**].

Anonymous Requests

Besides these login mechanisms, there is the concept of anonymous requests. When enabled, "requests that are not rejected by other configured authentication methods are treated as anonymous requests" [**the_linux_foundation_authenticating_2024**]. As such, a user would be able to query the API without identifying himself which, if proper RBAC configurations are not employed, could be quite damming to an organization. As such, it is advised that anonymous requests are disabled by using the `-anonymous-auth=false` on the API server. If it needs to be enabled due to organizational needs, it's paramount to ensure that the permissions of unauthenticated users are as restrictive as possible.

5.2.2 Admission Controllers

When considering interactions with the Kubernetes API, there are three actions that take place: authentication, authorization, and admission. This is clearly illustrated in Figure 5.2.2. It's worth highlighting that both human users, as well as Pods, go through the same access control flow, further showcasing the versatility of Kubernetes authentication and authorization methods.

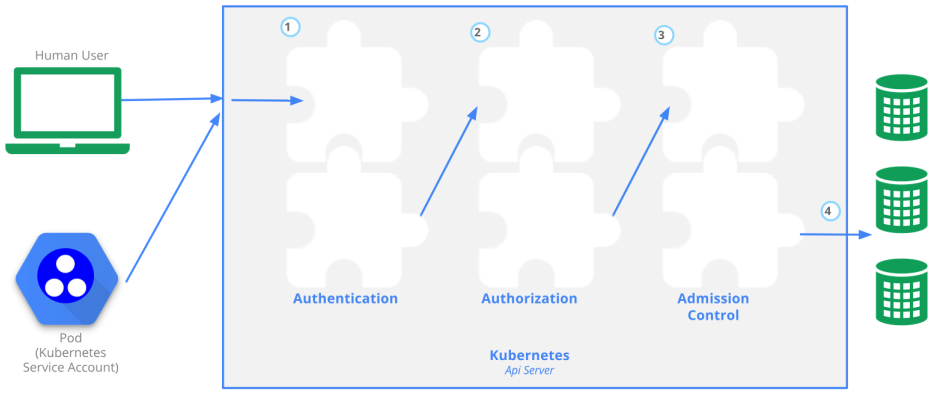

Figure 5.1: Access Control Overview [**the_linux_foundation_controlling_2023**]

The first one has been explored in 5.2.1, covering the different methods supported by Kubernetes, as well as recommendations. Considering authorization, this will be covered in 5.4. So, we must now cover the admission phase, as it's the last in the process, as well as a crucial part of it.

In Kubernetes, the admission process is handled by Admission Controllers. These are

"software modules that can modify or reject requests" [**the_linux_foundation_controlling_2023**]. They do so by intercepting requests to the API before the persistence of resources [**the_linux_foundation_admi**
Note that the persistence of resources is being specifically mentioned, excluding other types of operations. This is because admission controllers only operate on "requests that create, delete, or modify objects" [**the_linux_foundation_admission_2024**]. As such, read operations (the `get`, `watch` or `list` verbs) must be strictly covered by RBAC policies.

Admission controllers may be validating, mutating, or both. With validating admission controllers, the resources are validated against the set of modules to ensure that they comply with the rules imposed by them, and are only persisted in the cluster if they pass all controllers. In the case of mutating controllers, however, they may modify the data for the resource that's being modified [**the_linux_foundation_admission_2024**].

These modules are enabled at the API level, as we've seen. This is done by using the `-enable-admission-plugins` flag with the controllers that we want to enable. As such, to see what admission controllers are currently enabled in the cluster, we can investigate the manifest of the Pod that runs the API server in the cluster. An example of this is showcased in Figure 5.2.2.

```
luis@luis-Zenbook-UM3402YA-UM3402YA:~$ kubectl get pods -A
NAMESPACE     NAME                                 READY   STATUS    RESTARTS   AGE
kube-system   coredns-7db6d8ff4d-8vqx9             1/1     Running   0          8m15s
kube-system   etcd-minikube                        1/1     Running   0          8m29s
kube-system   kube-apiserver-minikube              1/1     Running   0          8m29s
kube-system   kube-controller-manager-minikube     1/1     Running   0          8m29s
kube-system   kube-proxy-8kbx9                     1/1     Running   0          8m15s
kube-system   kube-scheduler-minikube              1/1     Running   0          8m29s
kube-system   storage-provisioner                  1/1     Running   0          8m27s
luis@luis-Zenbook-UM3402YA-UM3402YA:~$ kubectl get pod kube-apiserver-minikube -n kube
-system -o yaml | grep enable-admission-plugins
    - --enable-admission-plugins=NamespaceLifecycle,LimitRanger,ServiceAccount,Default
StorageClass,DefaultTolerationSeconds,NodeRestriction,MutatingAdmissionWebhook,Validat
ingAdmissionWebhook,ResourceQuota
luis@luis-Zenbook-UM3402YA-UM3402YA:~$
```

Figure 5.2: Enabled admission controllers in testing instance

By consulting the Kubernetes documentation in [**the_linux_foundation_admission_2024**], we're able to confirm that the default admission controllers in version 1.31 are the ones shown in 5.2.

Listing 5.2: Default Admission Controllers

```
CertificateApproval, CertificateSigning,
    CertificateSubjectRestriction, DefaultIngressClass,
    DefaultStorageClass, DefaultTolerationSeconds, LimitRanger,
    MutatingAdmissionWebhook, NamespaceLifecycle,
```

```
PersistentVolumeClaimResize, PodSecurity, Priority,
ResourceQuota, RuntimeClass, ServiceAccount,
StorageObjectInUseProtection, TaintNodesByCondition,
ValidatingAdmissionPolicy, ValidatingAdmissionWebhook
```

Thus, we're able to confirm that our testing instance is missing some of the defaults, which should be enabled. To enable them, we can simply edit the `kube-apiserver` manifest, adding the necessary controllers to the flag shown in Figure 5.2.2.

The available Admission Controller plugins vary vastly in functionality, and cover the validation and mutation of requests "mostly from the resource management and security points of view" [**sayfan_mastering_2020**]. Our recommendation is for practitioners to:

1. Validate what Admission Controllers are currently enabled in the cluster;
2. Investigate what plugins can be useful in their specific organizational context. This can be done by consulting the documentation of each controller's function [2], and evaluating if it responds to any organizational needs.

Besides these recommendations, we must also consider the ResourceQuota and LimitRanges admission controllers separately, as their implementation is quite helpful in ensuring the proper deployment of applications in a Kubernetes cluster.

ResourceQuotas and LimitRanges

Kubernetes offers the ResourceQuota and LimitRanges admission controllers by default, which can be used to restrict resource usage for namespaces, nodes or Pods [**national_security_agency_kube**

Resource quotas can be used to limit the aggregate resource consumption per namespace [**the_linux_foundation_resource_2024**]. With them, practitioners can manage their computational resources used by their clusters' namespaces more effectively. For instance, they can be used to limit the total amount of CPU and memory requested and/or used by each namespace.

It's worth noting the difference between the requested and the used values, as an administrator must be aware of this to properly define the resource quotas for the cluster. Considering the CPU and memory attributes, the requested value represents the resources the Pod specifies at the time of its initialization. This does not mean that the Pod will automatically use the requested amount of CPU and memory, but that it will require them during the execution. As an example, a Pod can request 1Gi of RAM, but only consume 100Mi while it's executing. However, if the memory usage has any spike

[2] Presented in this documentation link

during the Pod's operation - for instance, due to a spike in traffic - it will be able to use more of it, since that initial request has been allocated to it.

Further analyzing this example scenario, if this memory spike were to increase further, it could hinder the operation of the cluster as a whole, since it's exhausting the resources. This is where the concept of limits is useful. If a limit of memory is defined in this case, the pod will be able to use more that it has requested - as we've explained prior - but not more than the defined limit.

Having this context established, we assert as a good practice to define resource quotas which impose maximum values not only on the compute power requested for a namespace, but also on the limits which said namespace will not be able to exceed. A practical example of a ResourceQuota manifest matching this description is presented in 5.3.

Listing 5.3: ResourceQuota example

```
apiVersion: v1
kind: ResourceQuota
metadata:
  name: example-cpu-mem-resourcequota
spec:
  hard:
    requests.cpu: "1"
        requests.memory: 1Gi
        limits.cpu: "2"
        limits.memory: 2Gi
```

Whereas a ResourceQuota imposes a limit on the resources used by a namespace, the LimitRange policy imposes limits per Pod [**national_security_agency_kubernetes_2022**]. It's objective is to help adminstrators ensure "that a single object cannot monopolize all available resources within a namespace" [**the_linux_foundation_limit_2024**]. As such, it provides constraints to limit resource allocations per Pod or even container within each Pod, as well as storage requests per each `PersistentVolumeClaim`. In 5.4, we present an example of a LimitRange manifest which could be applied in a cluster.

Listing 5.4: ResourceQuota example

```
apiVersion: v1
kind: LimitRange
metadata:
  name: cpu-resource-constraint
  namespace: demo
spec:
```

```
limits:
- default:
    cpu: 500m
  defaultRequest:
    cpu: 500m
  max:
    cpu: "1"
  min:
    cpu: 100m
  type: Container
```

What the presented manifest does is:

1. Define the default CPU limit and request for containers in the "demo" namespace as 500m. What this means is that, if no request or limit is specified in the Pod, it will be set to 500m;

2. Define that the maximum CPU a container can request is 1;

3. Define that the minimum CPU a container can request is 100m.

It's important to note that the LimitRange admission controller "does not check the consistency of the default values it applies" [**the_linux_foundation_limit_2024**]. This means that the values which are applied could generate a conflict with the container that's being started. For instance, and still considering the LimitRange in 5.4, if we were to start a Pod that specifies a container which requests 700m of CPU, this policy wouldn't alter this request, but would set the limit to 500m. Thus, the Pod wouldn't be scheduled, since the request can never be bigger than the limit.

We advise cluster administrators to set `ResourceQuotas` and `LimitRanges` according to the needs of their applications. While not strictly a security concern, it could mitigate potential issues in the cluster, which could originate some exposure to attacks.

5.2.3 API Server Security Flags

Throughout this chapter, we've already explored some API configurations that can be enabled to harden the security posture of a Kubernetes cluster. Besides those, there are several others which we find relevant to systematize in the present document. That is, therefore, the goal that this subsection aims to achieve.

In 2020, the vulnerability CVE-2020-8558 [**national_institute_of_standards_and_technology_nvd_2020**] was reported. It states that there was a defect in the kubelet and kube-proxy components which allowed attackers on the local network of the cluster, or in malicious

pods with specific capabilities running in it, to exploit the API server. This was because they were able to access services bound to the localhost address from the local network, which should not be possible. This highlighted the importance of ensuring that the flag `-secure-port` is configured, so that the API listens on a secure port [**the_linux_foundation_kube-apiserver_2024**].

Also related to the `-secure-port` flag, we advise administrators to also set the `-bind-address` configuration, which sets the IP address on which the API will listen to for requests [**the_linux_foundation_kube-apiserver_2024**]. This should be set to an address on the local network, ensuring that the API is not exposed on a server's public IP. Keep in mind that, if this is implemented, a VPN or reverse proxy will be required for administrators to communicate with the API.

The Kubernetes API also allows administrators to configure audit policies [**the_linux_foundation_auditing_**
These record "the chronological set of records documenting the sequence of actions in a cluster" [**the_linux_foundation_auditing_2024**], logging activities that are performed by users, by applications that interact with the API, and by the control plane itself. Similarly to the `EncryptionConfiguration` shown in 5.1, this feature requires the definition of a manifest, which is then referenced on the Kubernetes API configuration. In 5.5, we present an example audit policy, which we'll explore.

Listing 5.5: ResourceQuota example

```yaml
apiVersion: audit.k8s.io/v1
kind: Policy
rules:
  - level: Metadata
    verbs: ["create"]
    namespaces: ["demo"]
    resources:
      - group: ""
    resources: ["pods"]

  - level: RequestResponse
    resources:
      - group: ""
        resources: ["namespaces"]
```

In this example, we're logging two types of activities:

1. All creations of Pods in the "demo" namespace;

2. All activities related to namespaces in the cluster, including creation and deletion.

We can also see different values in the "level" key of each entry in the "rules" list. There are four possible values [**the_linux_foundation_auditing_2024**]:

- None: Does not log the events that match the rule;
- Metadata: Logs the event that matches the rule with the respective metadata, but doesn't include the request or response body;
- Request: Logs the events with request metadata and body, but does not log the response body;
- RequestResponse: Logs events with the request and response's metadata and body.

So, for the creation of namespaces, the policy in 5.5 will log both the request and response's metadata and body, whereas for the creation of pods, only the metadata will be recorded.

Now that the `Policy` manifest has been created, it must be referenced in the API configuration so that it can be used by the cluster. This is done by using the `audit-policy-file` and `audit-log-path` fields when configuring the API, as shown in Figure 5.3. Bear in mind that, for this example, we're outputting the logs to the standard output, which is the logs of the `kube-apiserver` pod.

```
root@minikube:/etc/kubernetes/manifests# cat kube-apiserver.yaml
apiVersion: v1
kind: Pod
metadata:
  annotations:
    kubeadm.kubernetes.io/kube-apiserver.advertise-address.endpoint: 192.168.49.2:8443
  creationTimestamp: null
  labels:
    component: kube-apiserver
    tier: control-plane
  name: kube-apiserver
  namespace: kube-system
spec:
  containers:
  - command:
    - kube-apiserver
    - --advertise-address=192.168.49.2
    - --allow-privileged=true
    - --audit-log-path=-
    - --audit-policy-file=/etc/ssl/certs/audit-policy.yaml
```

Figure 5.3: Configuration of API server for audit logs

As an example, Figure 5.4 demonstrates the audit log of a response to a request which creates the "demo" namespace.

In these logs, we can see some very useful information, such as the IP address from where the request originated, the username of who performed it and their RBAC groups.

```
{
  "kind": "Event",
  "apiVersion": "audit.k8s.io/v1",
  "level": "RequestResponse",
  "auditID": "f1cde366-6d67-4e2a-9bb3-58db030ada10",
  "stage": "ResponseComplete",
  "requestURI": "/api/v1/namespaces?fieldManager=kubectl-create&fieldValidation=Strict",
  "verb": "create",
  "user": {
    "username": "minikube-user",
    "groups": [
      "system:masters",
      "system:authenticated"
    ]
  },
  "sourceIPs": [
    "192.168.49.1"
  ],
  "userAgent": "kubectl/v1.29.10 (linux/amd64) kubernetes/f0c1ea8",
  "objectRef": {
    "resource": "namespaces",
    "name": "demo",
    "apiVersion": "v1"
  },
  "responseStatus": {
    "metadata": {},
    "code": 201
  },
  "requestObject": {
    "kind": "Namespace",
    "apiVersion": "v1",
    "metadata": {
      "name": "demo",
      "creationTimestamp": null,
      "labels": {
        "kubernetes.io/metadata.name": "demo"
      }
    },
    "spec": {},
    "status": {
      "phase": "Active"
    }
  },
  "responseObject": {
    "kind": "Namespace",
    "apiVersion": "v1",
    "metadata": {
      "name": "demo",
      "uid": "c2c74101-afcd-4c26-948b-99197e879df1",
      "resourceVersion": "569",
      "creationTimestamp": "2024-12-14T18:36:03Z",
      "labels": {
        "kubernetes.io/metadata.name": "demo"
:
```


Figure 5.4: Audit logs of a namespace creation

However, we can also infer that logging events with the `RequestReponse` level might not be appropriate for every case, as there's a lot of information to go through. For comparison, we present a *Metadata*-level log of a pod created in the "demo" namespace - matching the audit rule - in Figure 5.5.

From this log, we can clearly extract that the `demo-audit` pod was successfully created in the `demo` namespace - due to the 201 response code. We can also see exactly when it was created, and by what user and IP address.

The importance of these audit policies now becomes clear, as the data contained within them can assist practitioners in keeping track of the actions that take place in the cluster. Another factor which adds to the usability of this feature is the configurable granularity of the logs that are captured. With this, administrators can fine-tune audit policies based on their organizational requirements, thus ensuring that relevant actions are captured with the necessary detail.

```
{
  "kind": "Event",
  "apiVersion": "audit.k8s.io/v1",
  "level": "Metadata",
  "auditID": "0a7ab59a-04b5-4cc3-933d-32638ff54726",
  "stage": "ResponseComplete",
  "requestURI": "/api/v1/namespaces/demo/pods?fieldManager=kubectl-run",
  "verb": "create",
  "user": {
    "username": "minikube-user",
    "groups": [
      "system:masters",
      "system:authenticated"
    ]
  },
  "sourceIPs": [
    "192.168.49.1"
  ],
  "userAgent": "kubectl/v1.29.10 (linux/amd64) kubernetes/f0c1ea8",
  "objectRef": {
    "resource": "pods",
    "namespace": "demo",
    "name": "demo-audit",
    "apiVersion": "v1"
  },
  "responseStatus": {
    "metadata": {},
    "code": 201
  },
  "requestReceivedTimestamp": "2024-12-14T18:36:19.494674Z",
  "stageTimestamp": "2024-12-14T18:36:19.503382Z",
  "annotations": {
    "authorization.k8s.io/decision": "allow",
    "authorization.k8s.io/reason": "",
    "pod-security.kubernetes.io/enforce-policy": "privileged:latest"
  }
}
```


Figure 5.5: Audit logs of a pod creation

5.3 Network Policies

In order to control traffic flow in level 3 or 4 of the OSI layer, at IP address or port level, Network Policies rules should be implemented. These rules are meant to control communication within the cluster, between pods [**network_policies_kubernetes_from_kubernetes_io**].

Network Policies control traffic by filtering it and implementing network segregation. This allows the different components to be isolated and have a controlled access to other resources. This minimizes the attack surface and it plays a crucial role in maintaining integrity and confidentiality of applications.

By default, Kubernetes clusters have no policies set, which can cause security issues since all pods are allowed to communicate freely with each other. The administrator must explicity configure Network Policies to ensure a controlled and restrict environment within the cluster [**amaechi_learn_2023**].

To create any Network Policy, the cluster must use a network plugin, known as CNI. This is the actual controller that supports the Network Policy implementation, and without a controller the Network Policies will have no effect.

5.3.1 Policy Types

There are 2 policy types that must be implemented to ensure pod isolation: Ingress and Egress. While Ingress controls incoming traffic, Egress controls outgoing traffic.

As previously mentioned, no Network Policies are set by default. This means that all outbound and inbound traffic is allowed in a Pod. When this behavior is restricted by explicitly defining the rules, only connections originated by Pods specified in the Egress section of a NetworkPolicy are allowed. Additionally, reply traffic for those allowed connections is implicitly permitted. [**walker_kubernetes_2023**]

Similar behavior occurs for Ingress, that the only allowed connections into the pod are those specified in the Ingress section and the reply for these connections is implicitly permitted. Additionally, connections from the pod's node are also allowed.

Rules in both Ingress and Egress have an order, so the first rule defined is the first to be check, the second rules is validated after the first one, and so on.

An example from [**network_policies_kubernetes_from_kubernetes_io**] is used to demonstrate how to define Network Policy:

Listing 5.6: NetworkPolicies example

```
apiVersion: networking.k8s.io/v1
kind: NetworkPolicy
metadata:
  name: test-network-policy
  namespace: default
spec:
  podSelector:
        matchLabels:
          role: db
  policyTypes:
  - Ingress
  - Egress
  ingress:
  - from:
        - ipBlock:
          cidr: 172.17.0.0/16
          except:
          - 172.17.1.0/24
        - namespaceSelector:
          matchLabels:
          project: myproject
        - podSelector:
          matchLabels:
```

```
                  role: frontend
                ports:
                - protocol: TCP
                      port: 6379
        egress:
        - to:
                - ipBlock:
                  cidr: 10.0.0.0/24
                ports:
                - protocol: TCP
                  port: 5978
```

In this picture 5.6, the Network Policy defines that:

1. It is applied to all pods with role equal to `db` inside the `default` namespace
2. Both Ingress and Egress rules are set
3. Inbound traffic from range `172.17.0.0/16`, except IPs in range `172.17.1.0/24` is accepted. As well as traffic from Pods in namespaces that match label `project:myproject` and from Pods with label that matches `role:frontend`
4. Outbound traffic to range `10.0.0.0/24` with `TCP` protocol in port `5978` is allowed

5.3.2 Component Analysis

When defining a Network Policy, the following main components can be identified [**amaechi_learn_2023**]:

1. `podSelector`: defines to which pods the policy is applied to
2. `policyTypes`: indicates if an Ingress, Egress or both policies are applied. If no policyTypes are specified, the default is to set Ingress, and Egress if any egress rule is defined.
3. `ingress`: list of ingress rules
4. `egress`: list of egress rules

Ingress and Egress rules have a few different ways to be defined. As displayed in 5.6, the following components either for Ingress (using `from` block) or Egress (using

`to` block), can be used to define the elements that are allowed to establish a connection. The components are: `podSelector` that is a list of the Pods to be included, `namespaceSelector` that is a list of namespaces that Pods must be on, and `ipBlock` that contains a list of IPs or ranges that are allowed.

5.3.3 Best Practices

To maximize the security benefits of Network Policies, it is essential to adhere to established best practices. The following actions mentioned in [**amaechi_learn_2023**] and [**de_tender_kubernetes_2022**] must be implemented to ensure robust network isolation, prevent unauthorized access, and maintain secure communication within the Kubernetes cluster.

Use a default `deny-all` network policy. This ensures that only explicitly defined resources can communicate with the Pod **??**.

Listing 5.7: Deny All Default Policy example

```yaml
apiVersion: networking.k8s.io/v1
kind: NetworkPolicy
metadata:
  name: default-deny-ingress
spec:
  podSelector: {}
  policyTypes:
  - Ingress
```

Once this default policy is implemented, carefully define the required Ingress and Egress rules to ensure your application functions as intended. These rules should adhere to the principle of least privilege by allowing only the traffic that is explicitly necessary for the application to work as expected.

Group all Pods that need to communicate with each other using the `podSelector`. This practices allows for an easier network management, less prone to errors and more scalable. With this, a more secure environment is created since communication is restricted to only the intended Pods.

5.4 RBAC and Authentication

Role-Based Access Control (RBAC) and Authentication are essential mechanisms in Kubernetes to ensure secure and controlled access to resources. These mechanisms work

together to define and enforce access policies for users, applications, and processes [**rostami_role-based_2023**].

Authentication is the first step in the security model, where the identity of a user or service attempting to interact with the Kubernetes API is verified [**rostami_role-based_2023**]. Kubernetes supports various authentication methods, already detailed in 5.2.1.

However, authentication alone does not grant access to resources. This is where RBAC plays a critical role. RBAC determines what authenticated users or services are allowed to do within the cluster. It defines access policies by associating permissions with specific roles [**rostami_role-based_2023**].

5.4.1 Role and ClusterRole

An RBAC Role or ClusterRole contains the rules that represent a set of permissions. Permissions are added, and there is no deny rule; the permission is either granted or, by default, not granted [**the_linux_foundation_rbacauthorization_2024**].

A Role sets permissions within a specific namespace, so when the Role is created, a namespace must be specified. A ClusterRole is a non-namespaced resource. The resources have different names because a Kubernetes object either has a namespace or not, but never both [**the_linux_foundation_rbacauthorization_2024**].

ClusterRoles have different uses. It's possible to:

- Set permissions on namespaced resources and be granted access in individual namespace(s) or across all namespaces
- Set permissions on cluster-scoped resources

In summary, if you need to define a role within a namespace, the Role should be used. Otherwise, if you need to define a role cluster-wide, use the ClusterRole.

5.4.2 RoleBinding and ClusterRoleBinding

A RoleBinding grants the permissions defined in a Role to users or groups within a specific namespace. It links users (subjects) to a Role, which specifies what actions they can perform in that namespace. A ClusterRoleBinding, on the other hand, grants permissions cluster-wide, linking a user to a ClusterRole that applies across all namespaces [**the_linux_foundation_rbacauthorization_2024**].

A RoleBinding can reference a Role within the same namespace, or it can reference a ClusterRole, binding that ClusterRole to a specific namespace. To apply a Cluster-

Role across all namespaces present in the Cluster, ClusterRoleBinding should be used [**the_linux_foundation_rbacauthorization_2024**].

The name of both RoleBinding and ClusterRoleBinding must be a valid path segment name.

5.4.3 Implementation of RBAC

To implement RBAC effectively, it is essential to follow the principle of least privilege, granting only the necessary permissions to perform specific tasks. Creating granular roles for specific workloads can help minimize the risk of unauthorized access. Auditing access policies ensures that they remain effective and secure over time [**rostami_role-based_2023**].

When a request is made to the Kubernetes API server, it undergoes three steps. First, authentication, as mentioned previously, verifies the identity of the requester. Next, authorization checks whether the request complies with the RBAC policies. Finally, admission control applies additional checks to ensure compliance with cluster-wide rules. This layered approach ensures that only authorized requests gain access to resources and that all actions respect the defined policies [**the_linux_foundation_rbacauthorization_2024**].

6 Complementary Tooling

While Kubernetes provides robust built-in capabilities for container orchestration, there is a broad list of complementary tools that allow to extend these capabilities. These tools enhance areas such as advanced security features and comprehensive monitoring to strengthen cluster security. This section highlights several tools that offer security-enhancing features and explains how each contributes to improving cluster security.

6.1 Falco

Falco is an open-source security tool that can be used integrated with Kubernetes, focused on runtime security monitoring. It consumes Linux kernel syscalls, and applies pre-defined rules to these events and warns the developer when something unexpected happends [**falco_about_page**]. Falco usually sits accross a distributed infrastructure, providing real-time security monitoring and auditing for Kubernetes clusters.

By default, Falco includes a set of defined rules. However, these can be customized to align with the specific needs of the cluster and it typically address scenarios such as monitoring sensitive file access, detecting unauthorized privilege escalations, and more [**benhassan_advanced_2024**]. They can be defined as follows in 6.1:

Listing 6.1: Falco rule example

```
- rule: Outbound Connection to Unauthorized Port
  desc: Alert on outbound traffic to unauthorized ports
  condition: >
        outbound and not fd.sport in (http, https)
  output: "Suspicious connection detected (command=%proc.
      cmdline connection=%fd.name)"
  priority: WARNING
  tags: [network]
```

In this example 6.1, Falco will monitor outbound traffic to unauthorized ports. This rule is implemented to prevent suspicious activity or data exfiltration attempts.

Falco excels in identifying unexpected behaviors in a runtime environment. This characteristic enables its application in various use cases, such as:

- Detecting Unauthorized Network Connections
- Monitoring for Changes in Critical Configuration Files
- Identifying Suspicious Pod Executions
- Detecting Kubernetes API Misuse

6.2 OPA (Open Policy Agent)

OPA is a tool that enables the implementation of policies across the cluster. Policies are defined in the OPA engine, and whenever a service (e.g. a Pod) in the cluster queries OPA, it evaluates the request and provides a decision based on the defined policies [**bruner_better_2020**].

It is important to note that OPA itself does not enforce policies, just a decision. This decision is made based on the policies defined before deploying OPA. These policies are written in a specific language called Rego [**barua_securing_2023**].

Once the OPA is configured and deployed (usually in server mode on a container), the OPA engine intercepts the request and makes a decision based on the policies defined. After a decision is made, OPA returns the decision via API in a JSON format. With this decision, the Admission Controller acts on the decision, either blocking or accepting the request [**opa_doc_guides_nodate**].

Using the example provided in [**bruner_better_2020**], the policy displayed in 6.2 reject names that do not start with the letter 'a'.

Listing 6.2: OPA policy example

```
package startswitha

name := input.request.name

default allow = false

allow {
  startswith(name, 'a')
}
```

In this code 6.2, in the first line is defined the name of the policy's namespace. Then, it is assgined to variable `name` the value of key `name` from the payload. In the thrid instruction, it is defined the default value of this policy to `false` and then, the rule is defined. In the final block, the `allow` statement acts like a function, if the validation

is successfull, `allow` is automatically changed to true, otherwise the rule stops and `allow` is left unchanged.

In this example 6.2, if a request to OPA with payload `{"name": "apple"}` the request is allowed by OPA. Yet, if a request to OPA with payload `{"name": "rice"}` will not be allowed.

6.3 Trivy

When managing Kubernetes clusters, it is crucial to identify if the resources that are being use have any vulnerability. One of the best tools to solve this problem is Trivy. Trivy is a vulnerability scanner for containers and other resources that can detect vulnerabilities in the dependencies being used [**benhassan_deep_2024**].

Trivy differentiates between three elements when scanning a Kubernetes cluster and its associated resources [**trivy_doc_kubernetes_nodate**]:

- Cluster Infrastructure: Components like the API server, kubelet, and addons.
- Cluster Configuration: Kubernetes resources like Roles and ClusterRoles.
- Application Workloads: Containerized applications such as Nginx or PostgreSQL.

When scanning each of these elements, Trivy scans the container image separately from the Kubernetes manifest that contains all resources definitions.

For the container image and the cluster, both are scanned for vulnerabilities, misconfigurations and exposed secrets [**trivy_doc_kubernetes_nodate**].

It as many advantages such as comprehensive scanning since it scans containers, filesystems, and repositories for vulnerabilities. It is easy to use because it has a simple installation, good documentation and the output from each of the commands is easy to understand. Finally, it can be integrated in a CI/CD pipeline, enable the automation of the scanning process [**benhassan_deep_2024**].

In 6.3, it is presented the result of a Trivy scanning in summary of a Kubernetes cluster:

6.4 Vault (by HashiCorp)

Effective management of secrets in Kubernetes is crucial for maintaining the security and integrity of your applications. Although Kubernetes has built-in secrets management mechanisms, such as Kubernetes Secrets, it has some limitations [**muppeda_guide_vault_2024**].

```
trivy [main] ⚡ trivy k8s --report summary
204 / 204 [----------------------------------------------------------------------------------------------------------------------------------------------------------

Summary Report for kind-kind

Workload Assessment
```

Namespace	Resource	Vulnerabilities					Misconfigurations					Secrets				
		C	H	M	L	U	C	H	M	L	U	C	H	M	L	U
local-path-storage	Deployment/local-path-provisioner	4	32	10	2			1	3	9						

```
Severities: C=CRITICAL H=HIGH M=MEDIUM L=LOW U=UNKNOWN

Infra Assessment
```

Namespace	Resource	Vulnerabilities					Misconfigurations					Secrets				
		C	H	M	L	U	C	H	M	L	U	C	H	M	L	U
kube-system	Pod/etcd-kind-control-plane					6		2	4	7						
kube-system	Pod/kube-apiserver-kind-control-plane					5		2	5	17						
kube-system	Deployment/coredns		13	11	3			1	3	4						
kube-system	DaemonSet/kube-proxy	19	49	49	81	3		3	4	9						
kube-system	DaemonSet/kindnet	19	60	60	81	3		3	5	5						
kube-system	Pod/kube-scheduler-kind-control-plane					5		2	4	9						
kube-system	ConfigMap/extension-apiserver-authentication								1							
kube-system	Pod/kube-controller-manager-kind-control-plane					5		2	4	11						
kube-system	ControlPlaneComponents/k8s.io/apiserver			3												
	Node/kind-control-plane		4	7	2		1	4		1						

```
Severities: C=CRITICAL H=HIGH M=MEDIUM L=LOW U=UNKNOWN
```

Figure 6.1: Trivy Summary Report [**trivy_doc_kubernetes_nodate**]

Vault is a tool desgined to solve these problems. It provides a robust way to securely store and manage sensitive information [**muppeda_guide_vault_2024**]. It provides multiple key features that ensure this security are [**vault_doc_introduction_nodate**]:

- Secure Secret Storage: Arbitrary key/value secrets can be stored in Vault. Before written them to persistent storage, these secrets are encrypted by Vault. If an attacker has access to the raw storage, it isn't enough to access your secrets.
- Dynamic Secrets: It has the ability to generate secrets on-demand for some systems. For example, when an application needs to access an S3 bucket, it requests credentials from Vault, which then creates an AWS keypair with the necessary permissions. Once these dynamic secrets are generated, Vault automatically revokes them when their lease expires.
- Data Encryption: Vault can encrypt and decrypt data without storing it. This allows teams to use Vault's encryption and descryption mechanism and not having to create a new one.
- Leasing and Renewal: All secrets in Vault have an expiration date. Once this secret's validation expires, Vault will automatically revoke that secret. This expiration date can later be renewed by clients via built-in renew APIs.
- Revocation: Vault offers native support for secret revocation, allowing it to revoke not just individual secrets, but entire groups of secrets. For instance, it can revoke all secrets accessed by a specific user or all secrets of a particular type.

Vault's core functionality revolves around the use of tokens, each linked to a specific client policy. These policies are path-based, detailing the actions and access permissions for each client within specified paths. Tokens can either be generated manually and allocated to clients, or clients can authenticate with Vault to receive a token. The diagram in 6.4 showcases Vault's fundamental workflow [**vault_doc_introduction_nodate**].

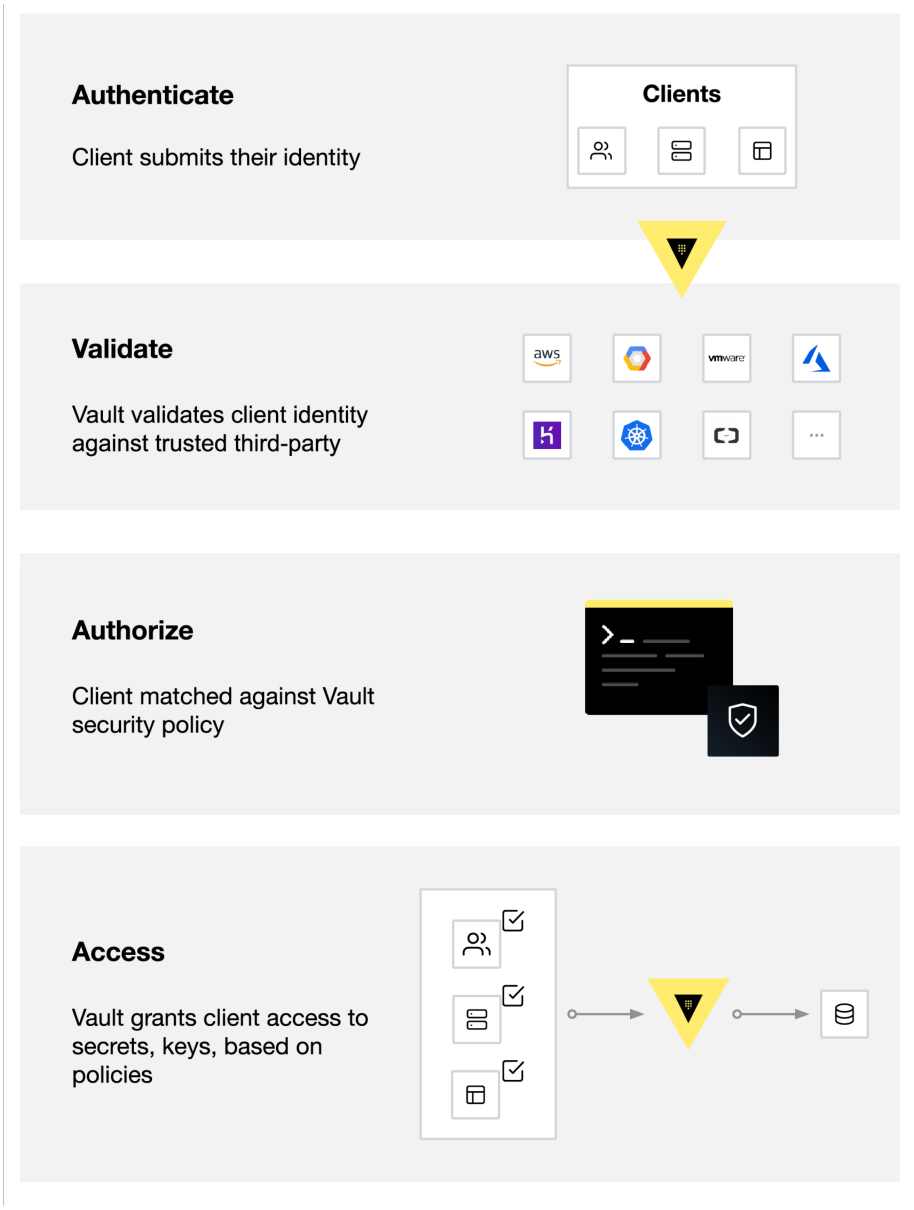

Figure 6.2: Vault Flow [**vault_doc_introduction_nodate**]

6.5 Kube-bench

Kube-bench is a tool that automates the validation of the security of the cluster against CIS Kubernetes Benchmark (set of the best security practices in Kubernetes). Helps to evaluate the cluster's configuration and Kubernetes components. After the check is done, the tool provides the administrator a detailed report with misconfigurations and vulnerabilities. Another of the advantages of using this tool is that it can be integrated in CI/CD pipelines, enabling this validation to be made continuously [**benhassan_depth_2024**] [**noauthor_aquasecurity_kube_bench_2024**].

One usual use case to use kube-bench is when there is a need to audit the security configuration of the master node of a cluster. After running the command to target the master node, an output similar to the one on 6.5 is displayed.

```
[INFO] 1 Master Node Security Configuration
[INFO] 1.1 API Server
[FAIL] 1.1.1 Ensure that the --allow-privileged argument is set to false (Scored)
[FAIL] 1.1.2 Ensure that the --anonymous-auth argument is set to false (Scored)
[PASS] 1.1.3 Ensure that the --basic-auth-file argument is not set (Scored)
[PASS] 1.1.4 Ensure that the --insecure-allow-any-token argument is not set (Scored)
[FAIL] 1.1.5 Ensure that the --kubelet-https argument is set to true (Scored)
[PASS] 1.1.6 Ensure that the --insecure-bind-address argument is not set (Scored)
[PASS] 1.1.7 Ensure that the --insecure-port argument is set to 0 (Scored)
[PASS] 1.1.8 Ensure that the --secure-port argument is not set to 0 (Scored)
[FAIL] 1.1.9 Ensure that the --profiling argument is set to false (Scored)
[FAIL] 1.1.10 Ensure that the --repair-malformed-updates argument is set to false (Scored)
[PASS] 1.1.11 Ensure that the admission control policy is not set to AlwaysAdmit (Scored)
[FAIL] 1.1.12 Ensure that the admission control policy is set to AlwaysPullImages (Scored)
[FAIL] 1.1.13 Ensure that the admission control policy is set to DenyEscalatingExec (Scored)
[FAIL] 1.1.14 Ensure that the admission control policy is set to SecurityContextDeny (Scored)
[PASS] 1.1.15 Ensure that the admission control policy is set to NamespaceLifecycle (Scored)
[FAIL] 1.1.16 Ensure that the --audit-log-path argument is set as appropriate (Scored)
[FAIL] 1.1.17 Ensure that the --audit-log-maxage argument is set to 30 or as appropriate (Scored)
[FAIL] 1.1.18 Ensure that the --audit-log-maxbackup argument is set to 10 or as appropriate (Scored)
[FAIL] 1.1.19 Ensure that the --audit-log-maxsize argument is set to 100 or as appropriate (Scored)
[PASS] 1.1.20 Ensure that the --authorization-mode argument is not set to AlwaysAllow (Scored)
[PASS] 1.1.21 Ensure that the --token-auth-file parameter is not set (Scored)
[FAIL] 1.1.22 Ensure that the --kubelet-certificate-authority argument is set as appropriate (Scored)
```

Figure 6.3: kube-bench output example [**noauthor_kube_bench_nodate**]

In this output, it is possible to observe that several critical checks have failed, indicating misconfigurations that could compromise the cluster's security. For example, this container allows privileged containers and insecure port configurations.

Although not displayed in the previous picture, for each failed test, kube-bench provides a description and remediation advice, similar to 6.5.

```
[FAIL] 2.1.2 Ensure that the --basic-auth-file argument is not set
[REMEDIATION] Edit the API server pod specification file /etc/kubernetes/manifests/kube-apiserver.yaml
on the master node and remove the --basic-auth-file argument.
```

Figure 6.4: kube-bench remediation output [**benhassan_depth_2024**]

7 Analysis and Best Practices

Throughout this document we've highlighted several configurations, as well as external tools, which assist practitioners in hardening the security of their Kubernetes clusters. The objective of this chapter is to consolidate these findings, and provide cluster administrators a set of recommendations and best practices for their cluster management operations.

Firstly, we highlight the importance of `etcd` encryption, notably through the usage of `EncryptionConfiguration` resources. Using `aescbc` is not recommended due to it being a weak provider. However, the `aesgcm` and `secretbox` providers are a security improvement from the default, clear-text way of storing secrets on `etcd`, with the downside of storing the encryption key in a local file-system. We note that this is the best option when using solely native configurations, but not the best solution altogether. As such, we urge practitioners to investigate the applicability of using an external KMS, and integrating it with the `kmsv2` encryption provider.

When considering the Kubernetes API Server, we strongly encourage administrators to tightly control access to it, using mechanisms such as firewalls and VPNs. Besides that base environment, we recommend the implementation of authentication and authorization measures, according to the least-privilege principle. The method used for this will vary across organizations, however, we must acknowledge the limitations of Kubernetes-native methods. Thus, we urge practitioners to implement authentication via an OIDC provider. For organizations without external user federation systems, such as Google SSO, we recommend the implementation of a Keycloak instance to manage the users within the organization, as it can then be integrated with the Kubernetes API for authentication purposes. Having this system, where every user has their own account, is also relevant for the auditing of actions, as well as forensic investigation in the case of an incident.

Having defined a method for authentication, we must now shift our focus towards authorization. This is where RBAC practices are relevant. We recommend that administrators define the necessary roles, according to the permissions necessary for each group in the organization. As organizational necessities shift with time, we encour-

age that a repeated procedure of auditing roles, as well as their `RoleBinding` and `ClusterRoleBinding` configurations, to ensure that they are up-to-date and only grant permissions which are strictly needed for a role.

Through our exploration of the topic, it's clear that networking is an important component on Kubernetes. As with the encryption of Secrets, there are no network security mechanisms between Pods by default. Thus, practitioners must implement them through the usage of `NetworkPolicies`. We strongly encourage the definition of a default deny-all policy, followed by policies allowing necessary traffic between pods and/or namespaces.

This study has highlighted that, although Kubernetes already has several security features, the usage of external tooling can complement these and further harden the security posture of a cluster. Practices such as runtime security, policy enforcement, vulnerability scanning and compliance benchmarks are paramount. In Chapter 6, we've presented tools for each of these security components, and recommended their implementation. However, it's worth noting that there are several technologies for these - and other - functionalities. A very useful resource for this is the CNCF Landscape [1], which presents technologies grouped by their respective area of operation, such as Observability, Security and Compliance.

Finally, we strongly encourage that cluster administrators consult the Kubernetes Security Checklist [2], as well as the OWASP Kubernetes Cheat Sheet[3]. These resources provide further guidance on how to harden Kubernetes environments, and enable practitioners to establish a robust security posture aligned with best practices and industry standards.

[1] Presented in: this link
[2] https://kubernetes.io/docs/concepts/security/security-checklist/
[3] https://cheatsheetseries.owasp.org/cheatsheets/Kubernetes_Security_Cheat_Sheet.html

8 Conclusions

The complexities and advantages of having Kubernetes for container orchestration in corporate environments have been thoroughly examined throughout this study. Key insights highlight the critical role of meticulous cluster component configuration and the adoption of security measures, such as robust data encryption protocols and effective secrets management.

Challenges, including the proper configuration of etcd and the protection of the API server, emphasize the necessity for a systematic and well-documented approach. This research reaffirms Kubernetes as a key technology for organizations striving for innovation, scalability, and enhanced application security.

Future research directions may include exploring multi-cloud integrations and complementary tools like service meshes to further enhance system flexibility and resilience. The findings presented here provide a valuable basis for advancing the study and practical application of Kubernetes in real-world scenarios.