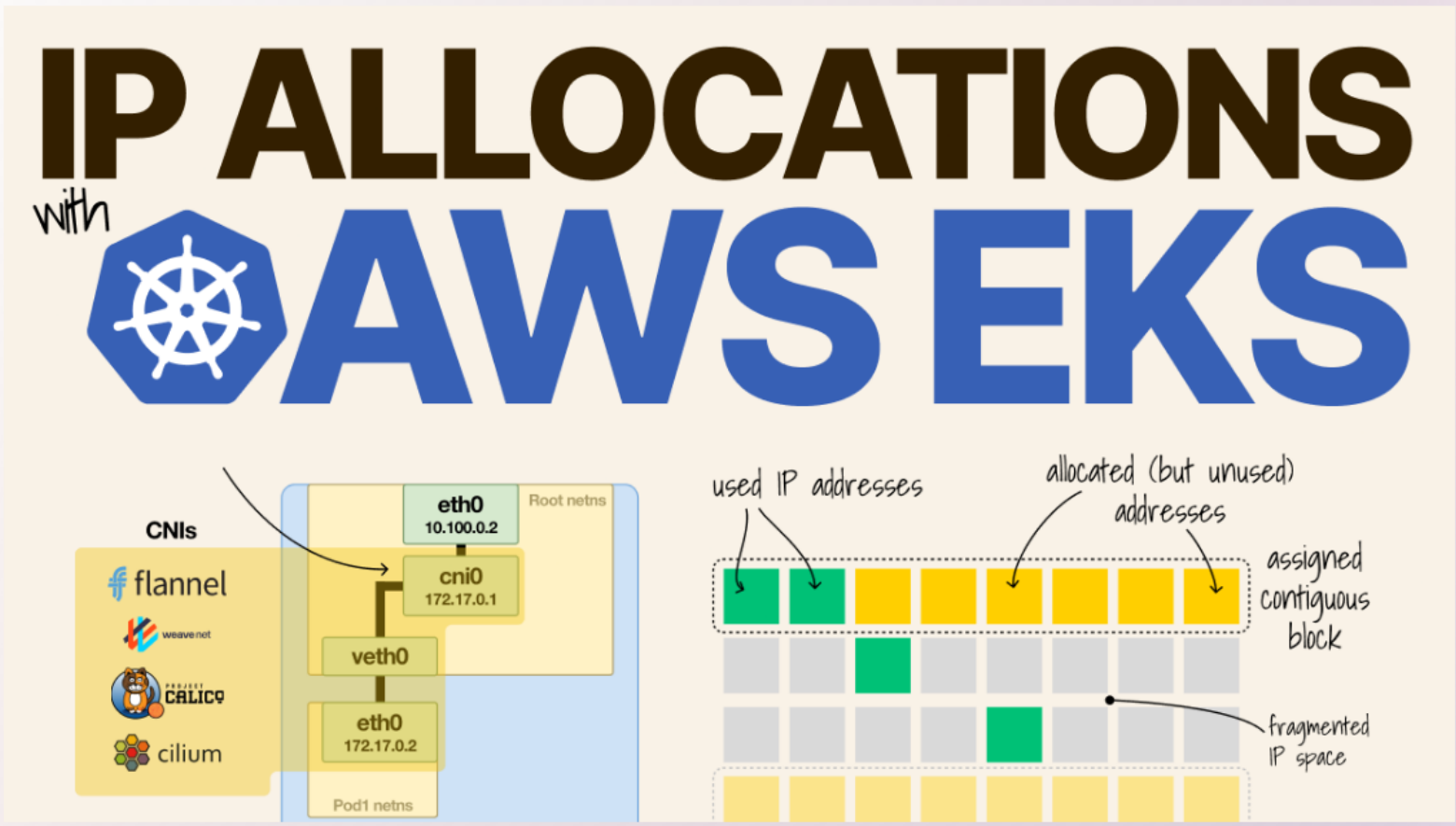
**Daniele Polencic — @danielepolencic@hachyderm.io**
@danielepolencic

When running an EKS cluster, you might face two issues:

- Running out of IP addresses assigned to pods.
- Low pod count per node (due to ENI limits).

Let me show you how you can fix those.




8:21 PM · 20 Mar, 2023

**15** replies **113** shares **491** likes

**Daniele Polencic — @danielepolencic@hachyderm.io**
@danielepolencic

1/

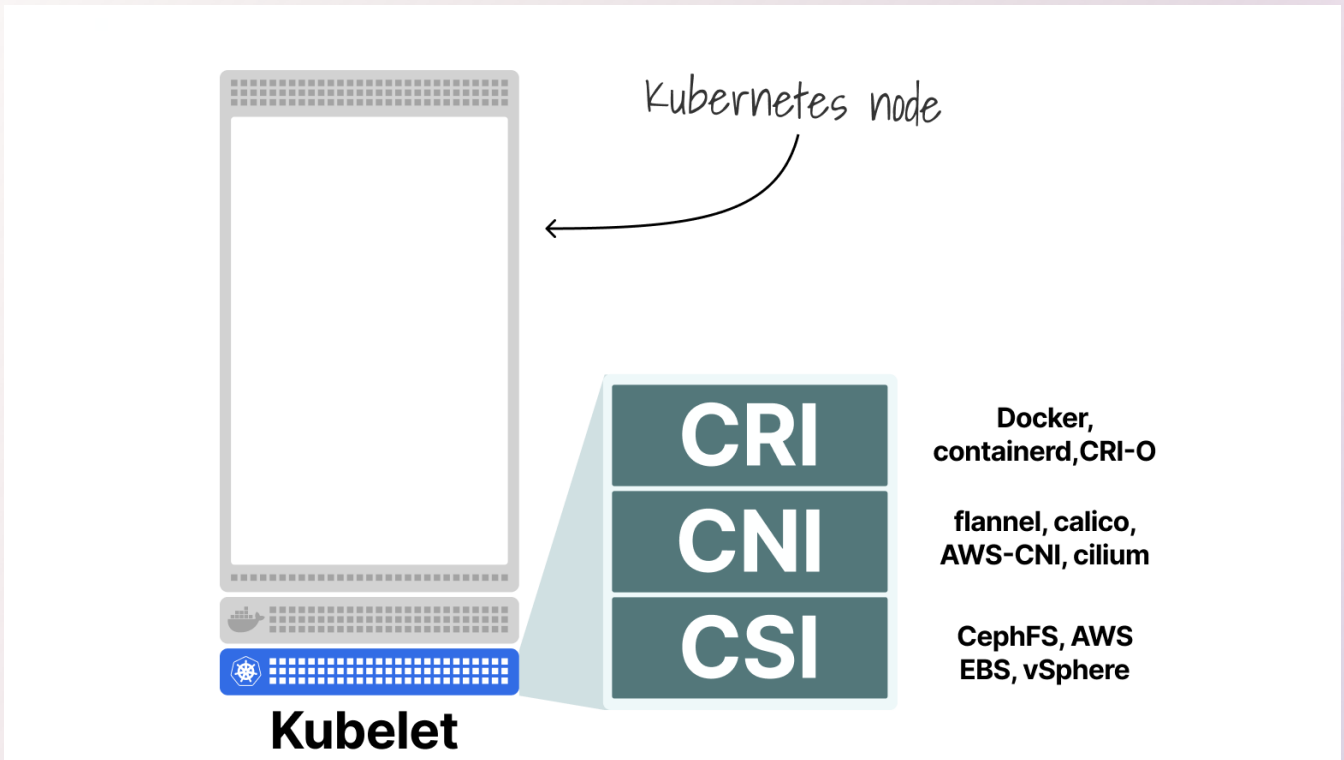
Before we start, here is some background on how intra-node networking works in Kubernetes

When a node is created, the kubelet delegates:

① Creating the container to the Container Runtime

② Attaching the container to the network to the CNI

③ Mounting volumes to the CSI




8:22 PM · 20 Mar, 2023

**1** reply **2** shares **12** likes

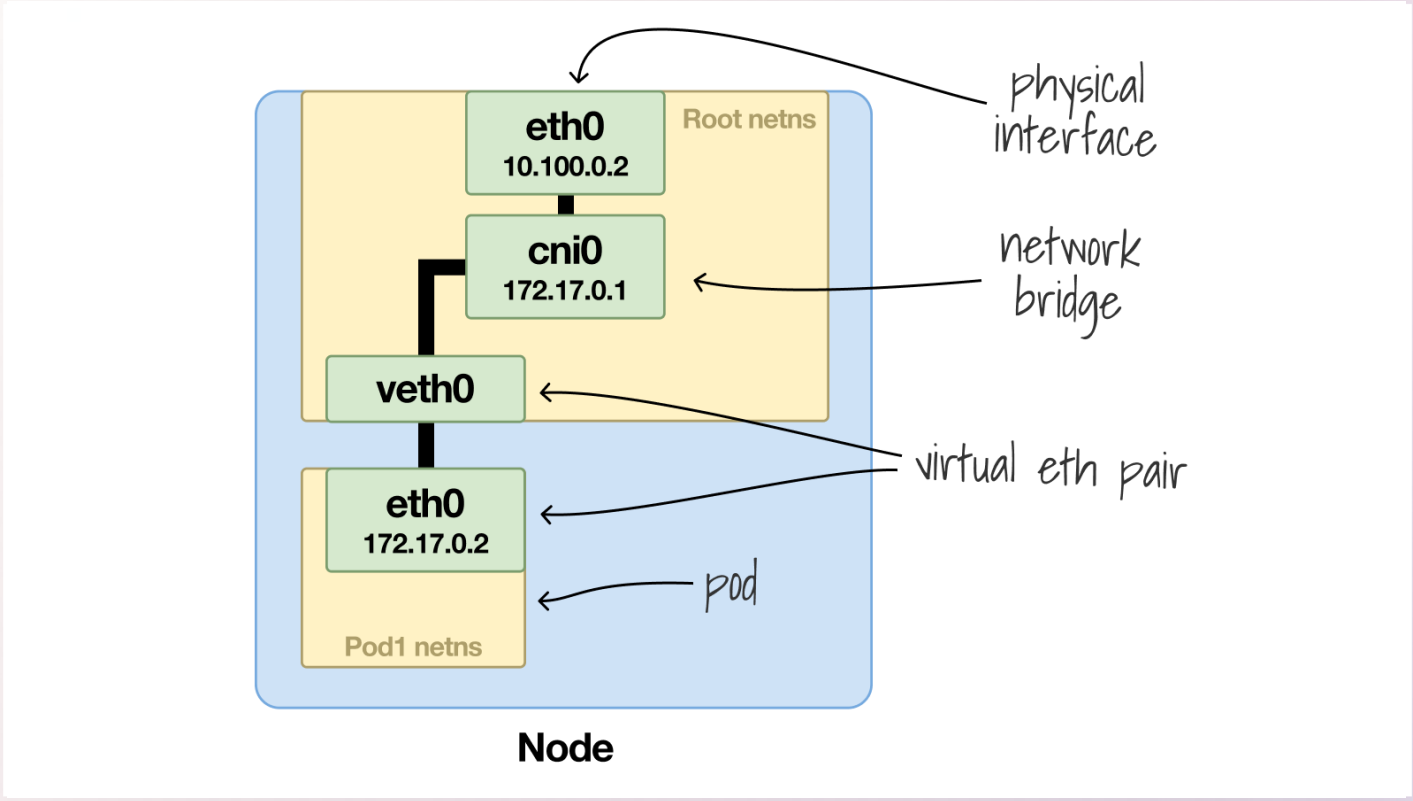
Daniele Polencic — @danielepolencic@hachyderm.io
@danielepolencic

2/

Let's focus on the CNI part

Each pod has its own isolated Linux network namespace and is attached to a bridge

The CNI is responsible for creating the bridge, assigning the IP and connecting veth0 to the cni0




8:22 PM · 20 Mar, 2023

**1** reply **9** likes

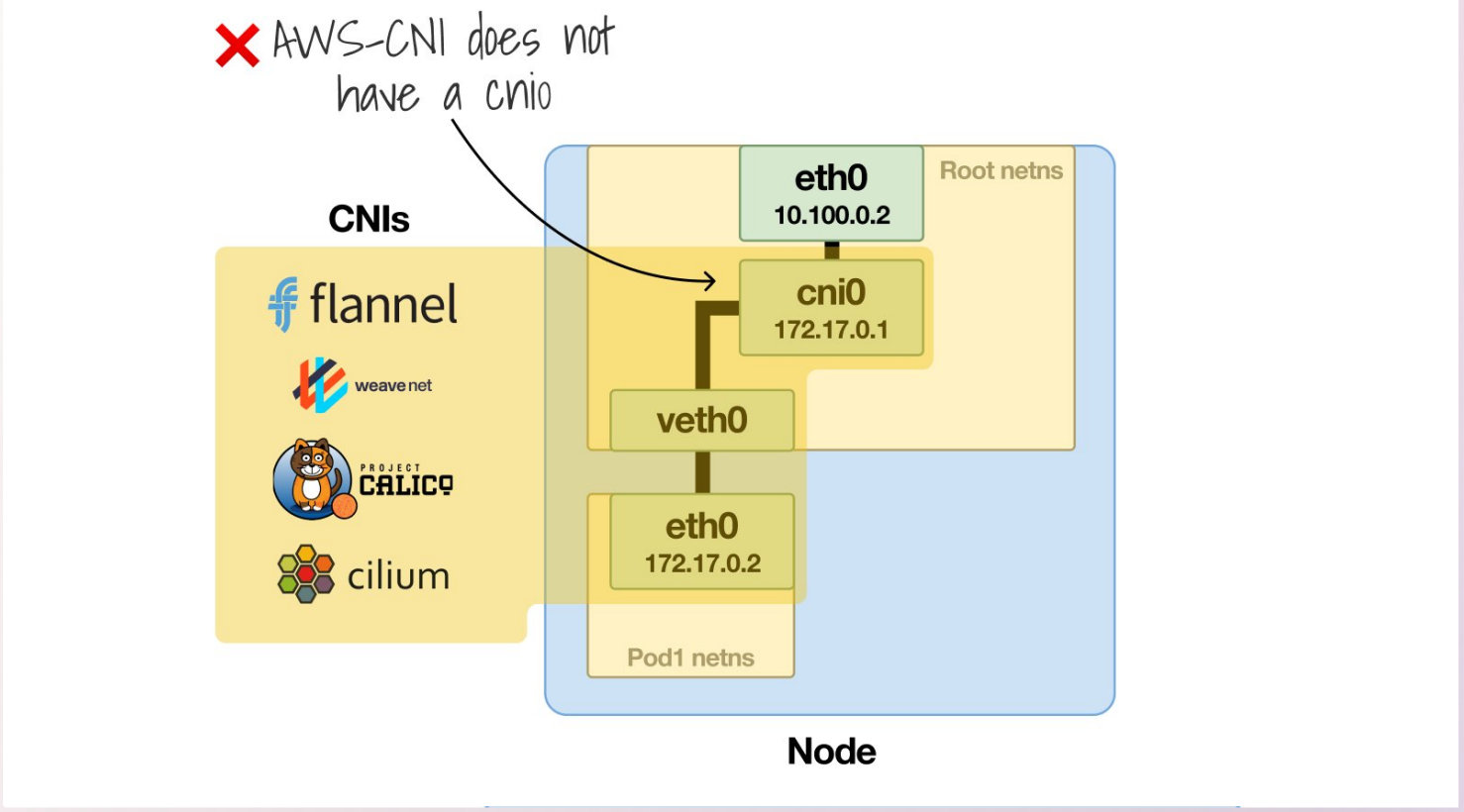
Daniele Polencic — @danielepolencic@hachyderm.io
@danielepolencic

3/

This usually happens, but different CNIs might use other means to connect the container to the network

As an example, there might not be a cni0 bridge

The AWS-CNI is an example of such a CNI




8:22 PM · 20 Mar, 2023

**1** reply **7** likes

**Daniele Polencic — @danielepolencic@hachyderm.io**
@danielepolencic

4/

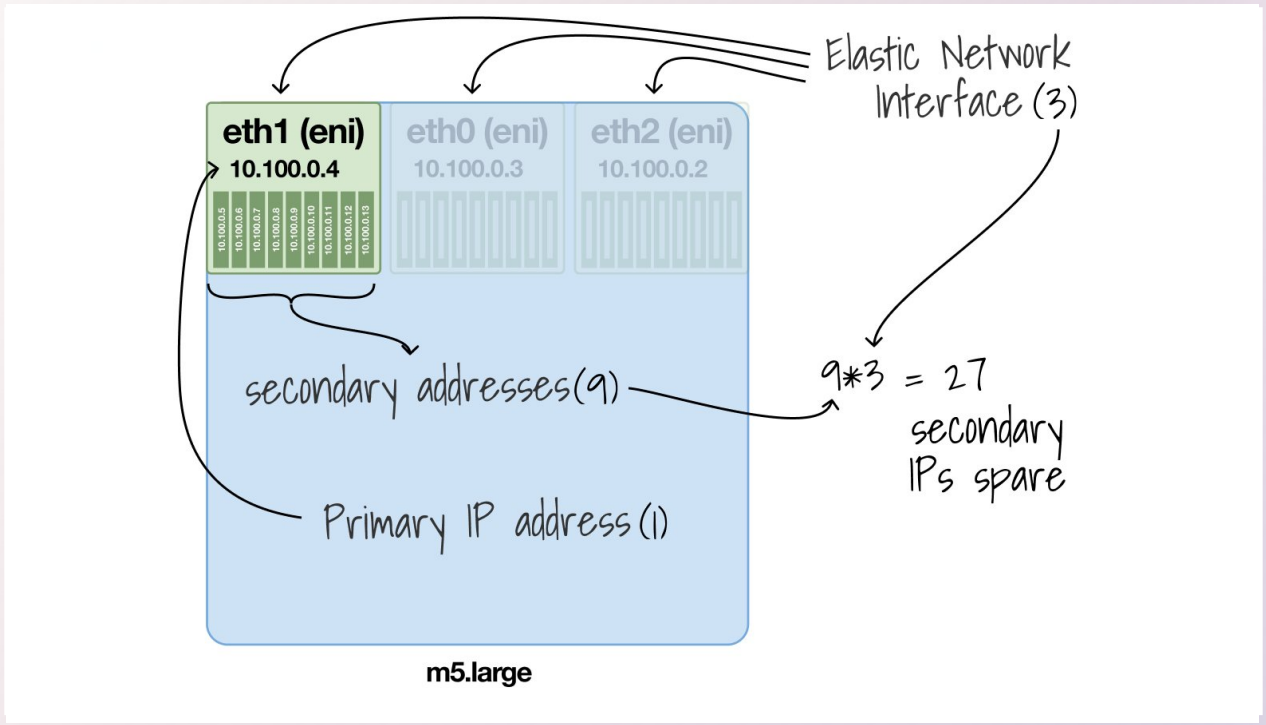
In AWS, each EC2 instance can have multiple network interfaces (ENIs)

You can assign a limited number of IPs to each ENI

For example, an m5.large can have up to 10 IPs for ENI

Of those 10 IPs, you have to assign one to the network interface

The rest you can give away




8:23 PM · 20 Mar, 2023

**2** replies **5** likes

**Daniele Polencic — @danielepolencic@hachyderm.io**
@danielepolencic

5/

Previously, you could use the extra IPs and assign them to Pods

But there was a big limit: the number of IP addresses

Let's have a look at an example

8:23 PM · 20 Mar, 2023

**1** reply **1** likes

**Daniele Polencic — @danielepolencic@hachyderm.io**
@danielepolencic

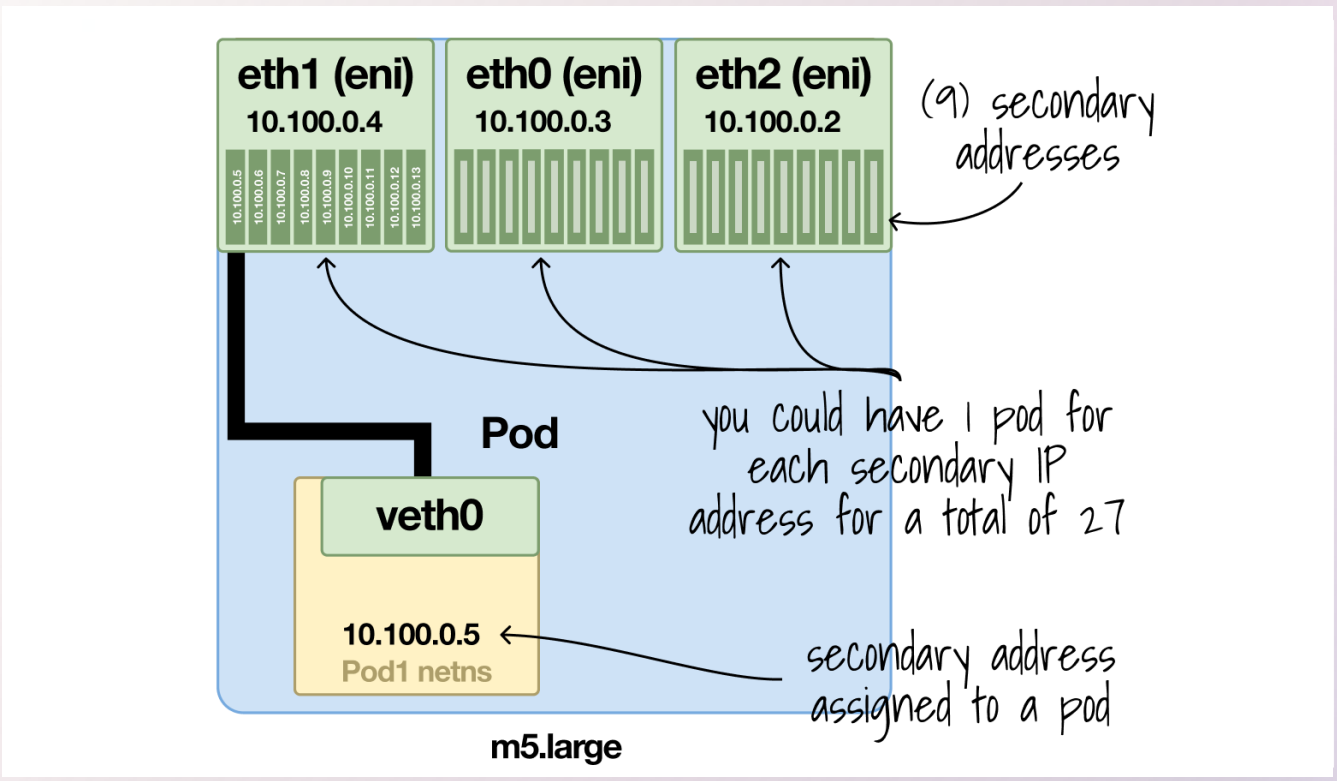
6/

With an m5.large, you have up to 3 ENIs with 10 IP private addresses each

Since one IP is reserved, you're left with 9 per ENI (or 27 in total)

That means that your m5.large could run up to 27 Pods

Not a lot




8:23 PM · 20 Mar, 2023

**1** reply **3** likes

**Daniele Polencic — @danielepolencic@hachyderm.io**
@danielepolencic

7/

But AWS released a change to EC2 that allows "prefixes" to be assigned to network interfaces

Prefixes what?!

In simple words, ENIs now support a range instead of a single IP address

8:23 PM · 20 Mar, 2023

**1** reply **5** likes

**Daniele Polencic — @danielepolencic@hachyderm.io**
@danielepolencic

8/

If before you could have 10 private IP addresses, now you can have 10 slots of IP addresses

And how big is the slot?

By default, 16 IP addresses

With 10 slots, you could have up to 160 IP addresses

That's a rather significant change!

Let's have a look at an example

8:23 PM · 20 Mar, 2023

**1** reply **3** likes

**Daniele Polencic — @danielepolencic@hachyderm.io**
@danielepolencic

9/

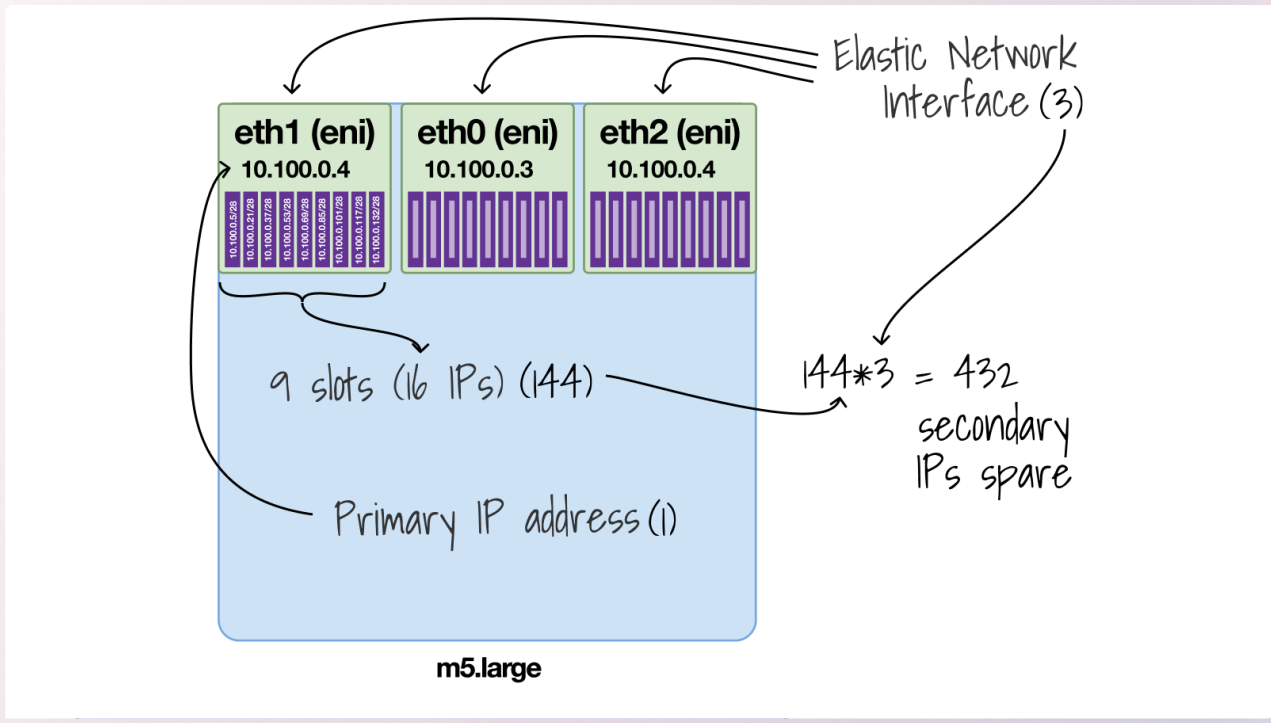
With an m5.large, you have 3 ENIs with 10 slots (or IPs) each

Since one IP is reserved for the ENI, you're left with 9 slots

Each slot is 16 IPs, so 9*16=144 IPs

Since there are 3 ENIs, 144x3=432 IPs

You can have up to 432 Pods now (vs 27 before)




8:24 PM · 20 Mar, 2023

**1** reply **3** likes

**Daniele Polencic — @danielepolencic@hachyderm.io**
@danielepolencic

10/

The AWS-CNI support slots and caps the max number of Pods to 110 or 250, so you won't be able to run 432 Pods on an m5.large

It's also worth pointing out that this is not enabled by default — not even in newer clusters

Perhaps because only nitro instances support it

8:24 PM · 20 Mar, 2023

**1** reply **2** likes

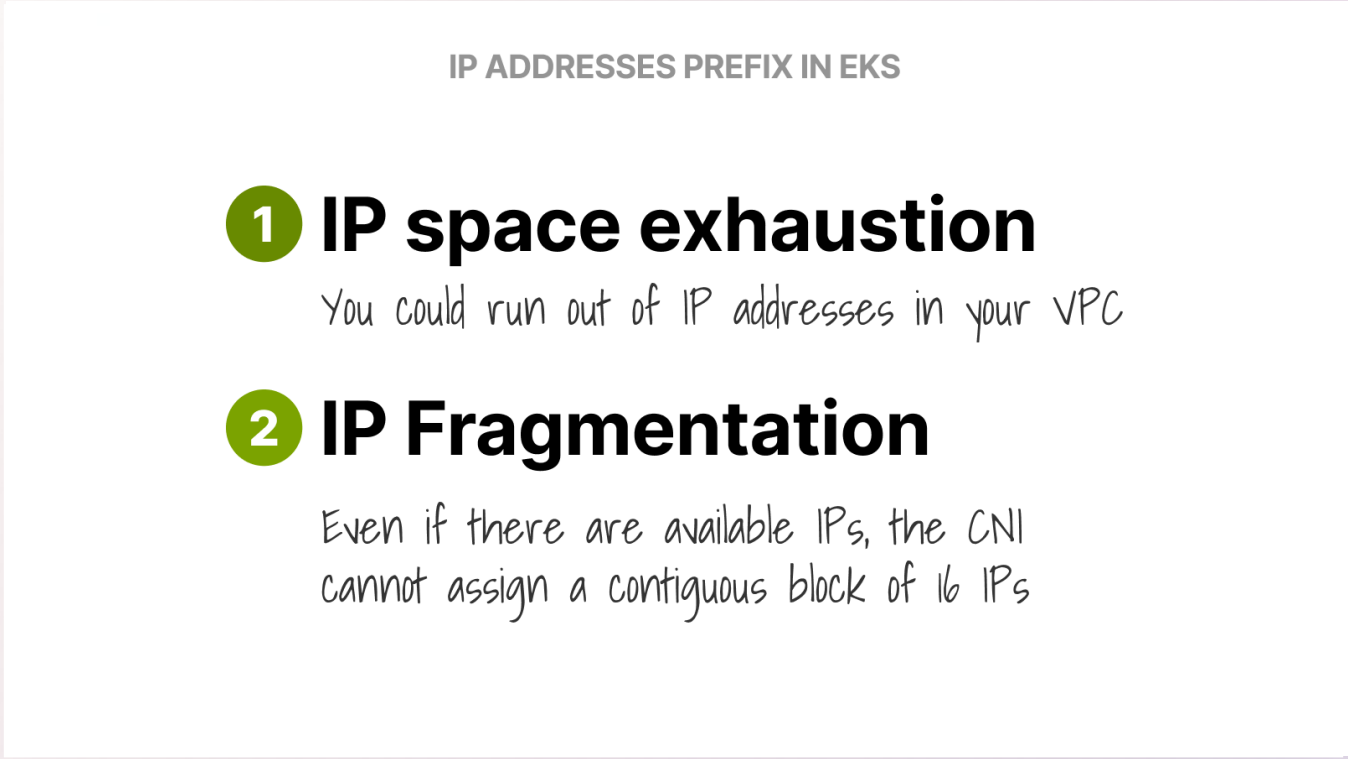
**Daniele Polencic — @danielepolencic@hachyderm.io**
@danielepolencic

11/

Assigning slots it's great until you realize that the CNI gives 16 IP addresses at once instead of only 1, which has the following implications:

- Quicker IP space exhaustion
- Fragmentation

Let's review those




8:24 PM · 20 Mar, 2023

**1** reply **5** likes

**Daniele Polencic — @danielepolencic@hachyderm.io**
@danielepolencic

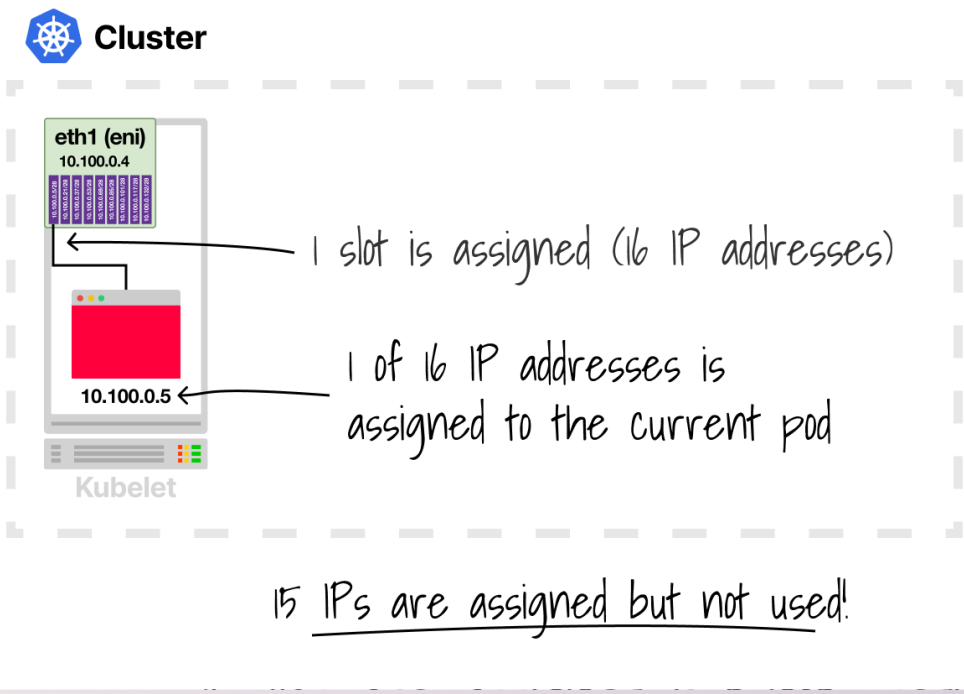
12/

A pod is scheduled to a node

The AWS-CNI allocates 1 slot (16 IPs), and the pod uses one

Now imagine having 5 nodes and a deployment with 5 replicas

What happens?




8:24 PM · 20 Mar, 2023

**1** reply **3** likes

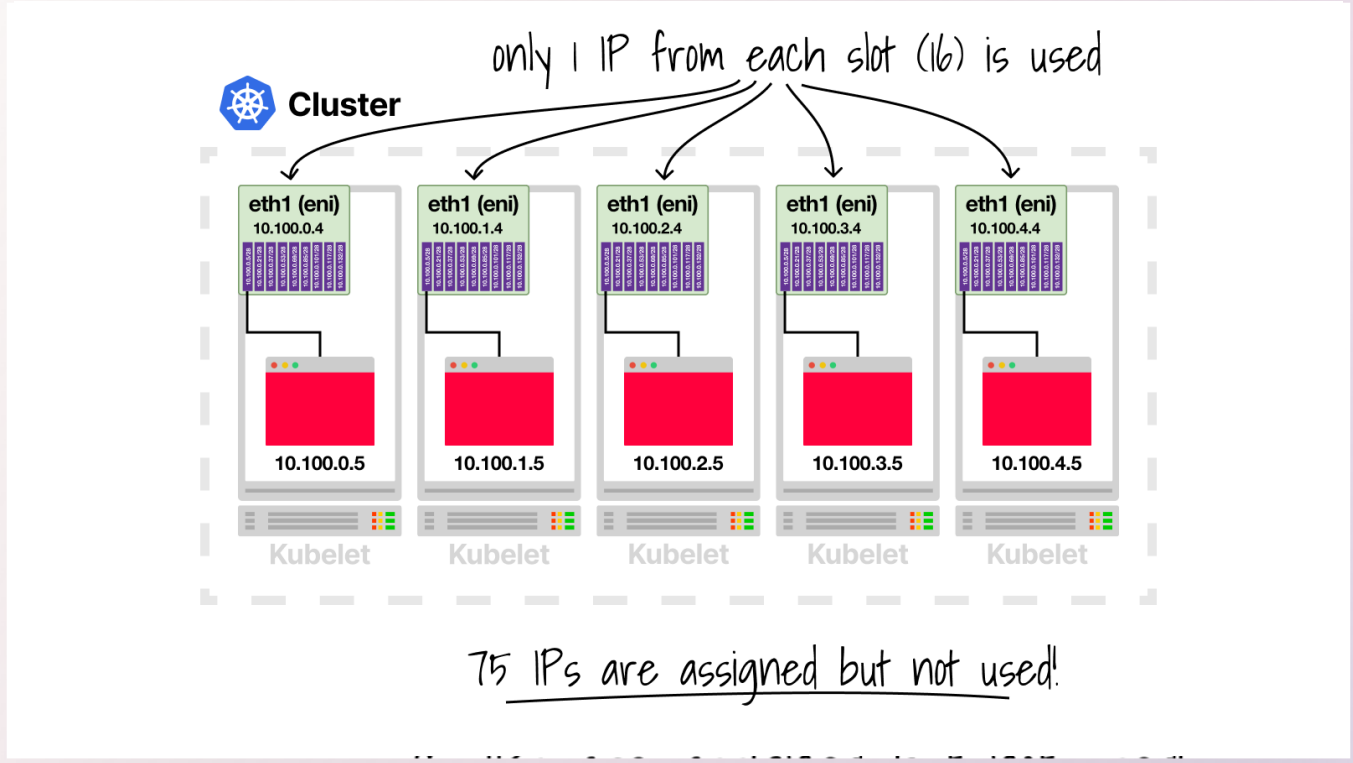
**Daniele Polencic — @danielepolencic@hachyderm.io**
@danielepolencic

13/

The Kubernetes scheduler prefers to spread the pods across the cluster

Likely, each node receives 1 pod, and the AWS-CNI allocates 1 slot (16 IPs)

You allocated 5*15=75 IPs from your network, but only 5 are used




8:25 PM · 20 Mar, 2023

**1** reply **4** likes

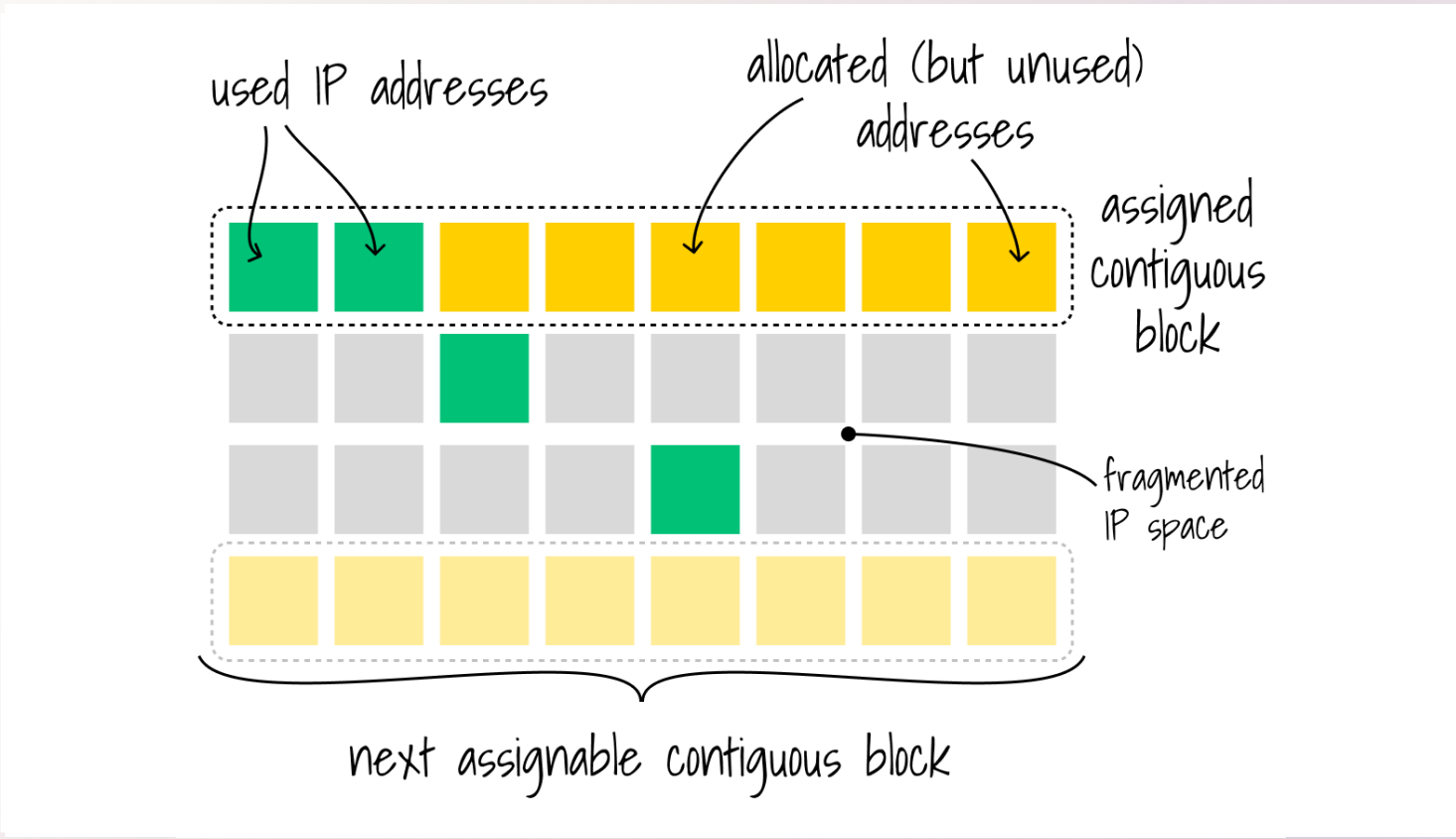
Daniele Polencic — @danielepolencic@hachyderm.io
@danielepolencic

14/

But there's more

Slots allocate a contiguous block of IP addresses

If a new IP is assigned (e.g. a node is created), you might have an issue with fragmentation




8:25 PM · 20 Mar, 2023

**1** reply **3** likes

**Daniele Polencic — @danielepolencic@hachyderm.io**
@danielepolencic

15/

How can you solve those?

- You can assign a secondary CIDR to EKS. aws.amazon.com/premiumsupport...

- You can reserve IP space within a subnet for exclusive use by slots. docs.aws.amazon.com/vpc/latest/use...

8:25 PM · 20 Mar, 2023

**1** reply **5** likes

**Daniele Polencic — @danielepolencic@hachyderm.io**
@danielepolencic

16/

Relevant links:

- docs.aws.amazon.com/AWSEC2/latest/...
- aws.amazon.com/blogs/containe...
- docs.aws.amazon.com/AWSEC2/latest/...

8:25 PM · 20 Mar, 2023

**1** reply **1** shares **7** likes

**Daniele Polencic — @danielepolencic@hachyderm.io**
@danielepolencic

17/

And finally, if you've enjoyed this thread, you might also like:

- The Kubernetes workshops that we run at Learnk8s
learnk8s.io/training

- This collection of past threads

- The Kubernetes newsletter I publish every week
learnk8s.io/learn-kubernet...

8:26 PM · 20 Mar, 2023

**7** likes