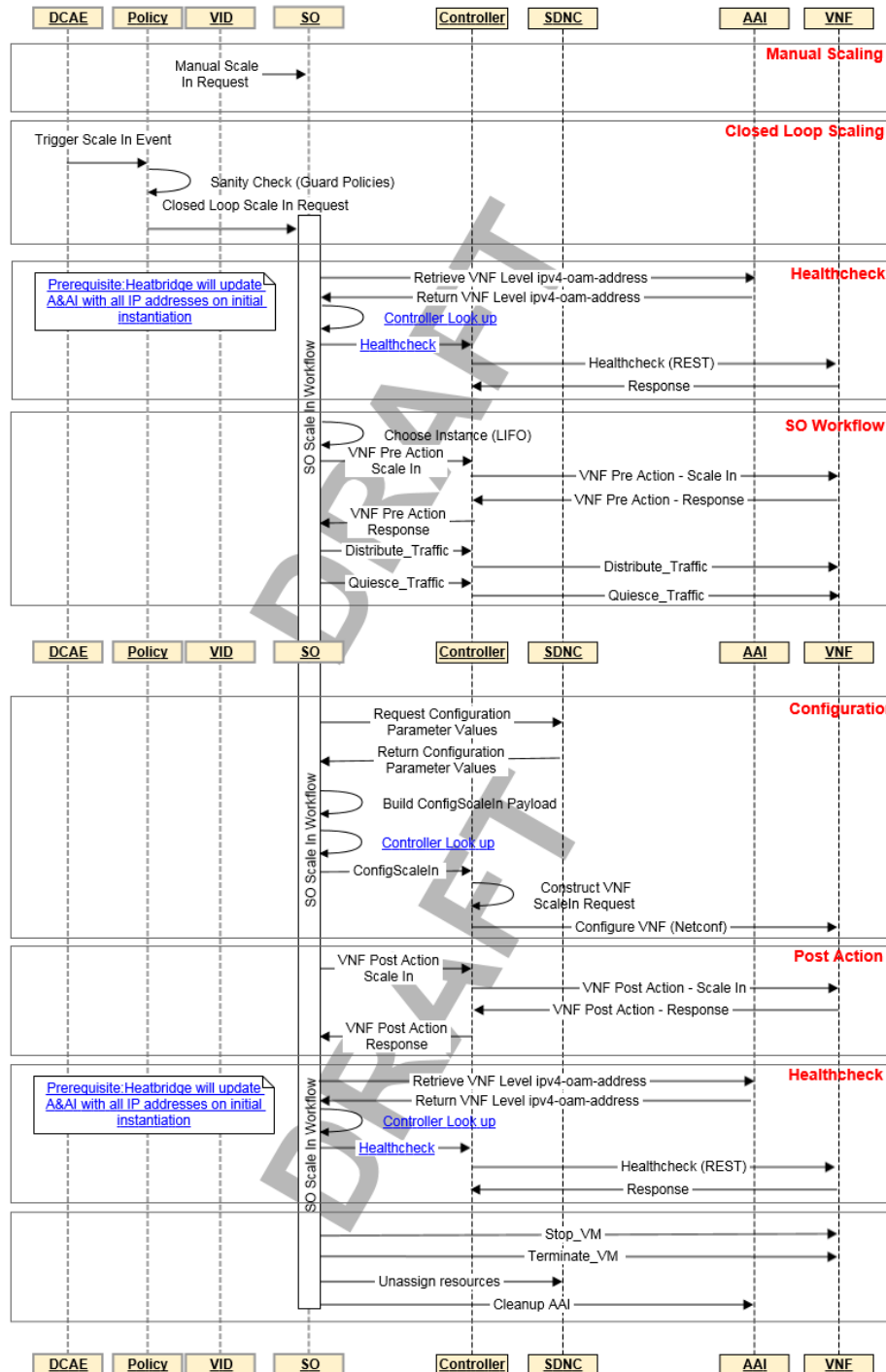


Scale In

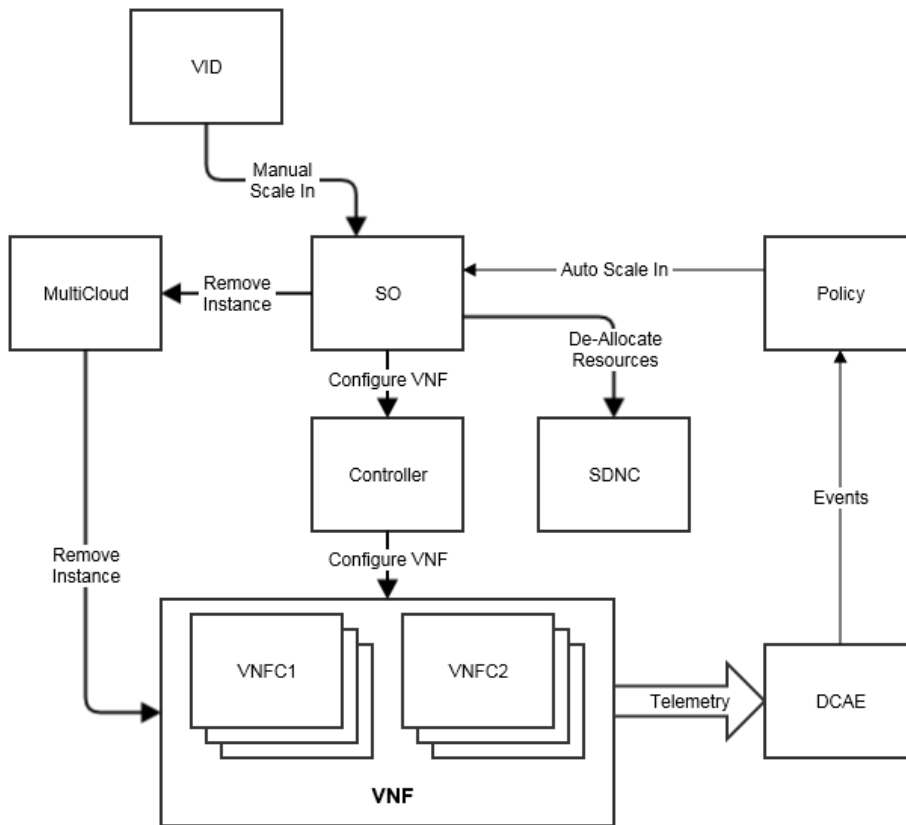
This page was used for initial draft solutions for Scale In. It has now been deprecated and is no longer up to date. The latest version can be found on the [Dublin Scaling Use Case wiki section](#).

Scale In Use Case Sequence Diagram - Run Time (DRAFT)



Issues to be resolved

- When should scaling be done?
 - In Dublin our primary reason for scaling will be focused on adjusting resources to the load. As load increases or decreases, the VNF will scale Out or In as appropriate. This can be done either manually or automatically.
 - Other reasons for scaling may include:
 - Upgrades
 - VM Moves
 - Meet with Operations to determine other reasons
- Which instance should be removed?
 - How should Policy be involved? (Long term: Policy should query an OOF microservice)
 - Last In First Out as a temporary solution? (Dublin: will need to query for all instances and then choose)
- We need a generic Pre-Action API with a scale in flag to trigger a scale in playbook. This will allow the VNF to do any VNF specific actions prior to a scale in action
- We need a generic Pre-Action API with a scale in flag to trigger a scale in playbook. This will allow the VNF to do any VNF specific actions prior to a scale in action
 - In onboarding package vendor needs to say whether or not they support a Pre and Post action
- How do we use Distribute_Traffic?



- a. Distribute_Traffic will move all new traffic
 - i. Will need to make Distribute_Traffic work at the VM /VNFC level
- b. Quiesce_traffic can be used for returning a response once all traffic is drained from the target instance
 - i. will need to make Quiesce_Traffic work on the VM /VNFC level
6. Need to build a ConfigScaleIn action and API for APPC and SDNC
7. Will we be able to use CDS for AutoConfiguration?
8. How do we determine which controller to use?
9. STOP VM before Termination
10. Scale In will release all resources. It is not guaranteed that the same resources will be reassigned on the next scale out.