



AM regular meeting

Lei Huang

Mar. 6, 2019

Topics

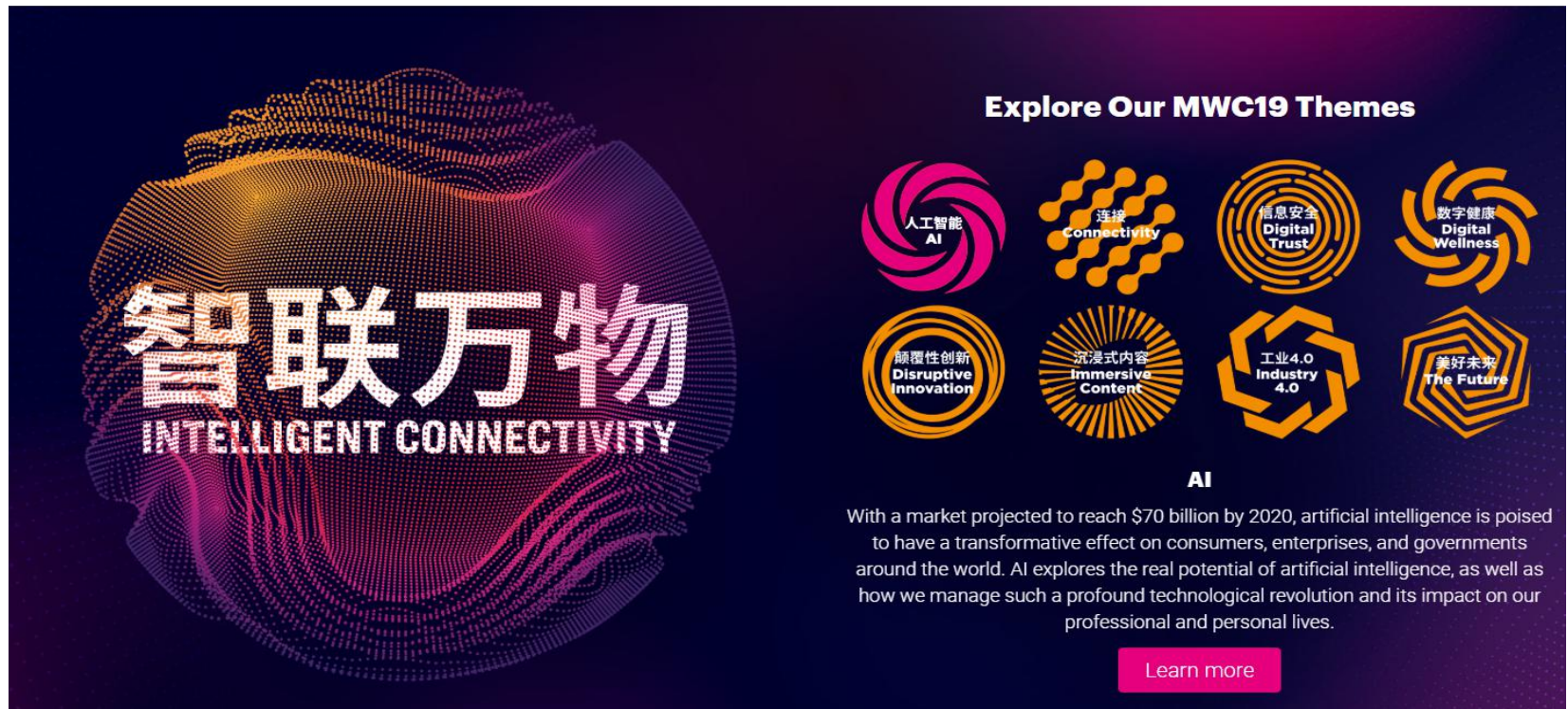
1. Requirements from CMCC in acceleration management
2. Discussion on AM workflow
3. BP of AM
4. Demo preparation
5. Call for volunteers

Requirements from CMCC in acceleration management

- Requirements from CMCC

We will have joint demo in MWC2019 -Barcelona mobile world congress Shanghai , 26-28 June, 2019.

<https://www.mwcshanghai.com/>



The graphic features a large, stylized circular logo on the left with the Chinese characters '智联万物' (Intelligent Connectivity) and the English text 'INTELLIGENT CONNECTIVITY' below it. The logo is composed of a grid of dots in shades of orange, yellow, and pink. To the right, under the heading 'Explore Our MWC19 Themes', there are eight circular icons representing different technology themes: AI (Artificial Intelligence), Connectivity, Digital Trust, Digital Wellness, Disruptive Innovation, Immersive Content, Industry 4.0, and The Future. Each icon contains its respective theme name in both Chinese and English.

Explore Our MWC19 Themes

人工智能 AI

连接 Connectivity

信息安全 Digital Trust

数字健康 Digital Wellness

颠覆性创新 Disruptive Innovation

沉浸式内容 Immersive Content

工业4.0 Industry 4.0

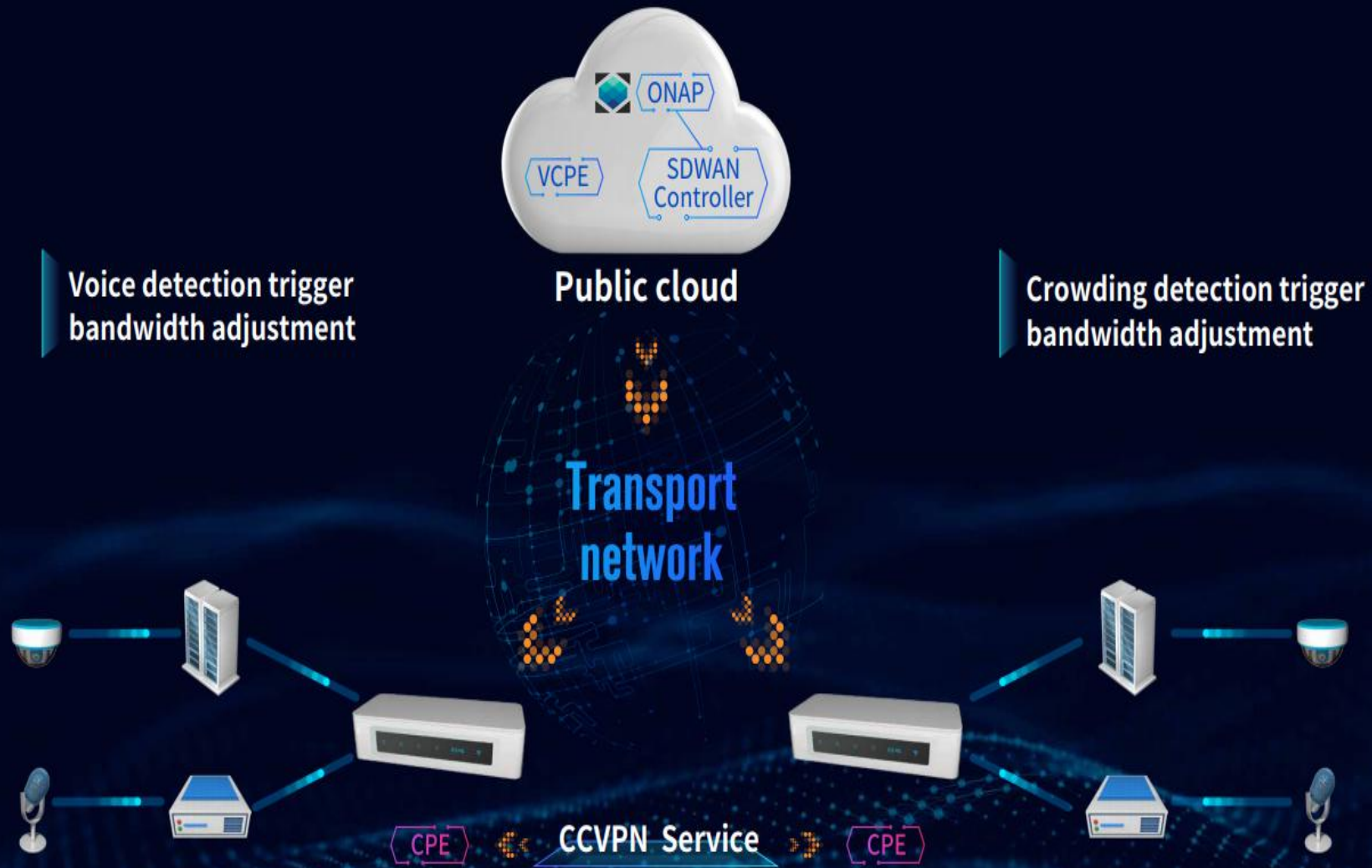
美好未来 The Future

AI

With a market projected to reach \$70 billion by 2020, artificial intelligence is poised to have a transformative effect on consumers, enterprises, and governments around the world. AI explores the real potential of artificial intelligence, as well as how we manage such a profound technological revolution and its impact on our professional and personal lives.

[Learn more](#)

AI-based network orchestration demonstration



Voice detection trigger bandwidth adjustment

Public cloud

Crowding detection trigger bandwidth adjustment

Transport network

CPE CCVPN Service CPE



Requirements from CMCC in acceleration management

- 1. CMCC demo currently needs acceleration technologies , such as GPU or FPGA in intelligent surveillance system. In the case of limited GPU resources, it is necessary to be able to realize the discovery process of proper acceleration resources and allocate resources reasonably in ONAP.
- 2. At present, the amount of video streaming data in intelligent security system is huge, by deploying the video stream through the edge cloud, and adding the monitoring and analysis mechanism in the edge cloud can alleviate the rate of large data flow processing of video stream.
- 3. When deploying GPU or FPGA resources, we need to increase the management capabilities of existing ONAP for GPU resources.
- 4. Users need to monitor and analyze the performance status of the GPU and the situation of the video stream
- 5. Allocate GPU resources appropriately to realize fast operation requirements of edge side computing equipment.
- 6. Deploy video stream in Edge cloud, allocate GPU resources reasonably, reduce the transmission pressure of video stream, and realize the optimal allocation of limited GPU resources

Discussion on AM workflow

- From users' perspective

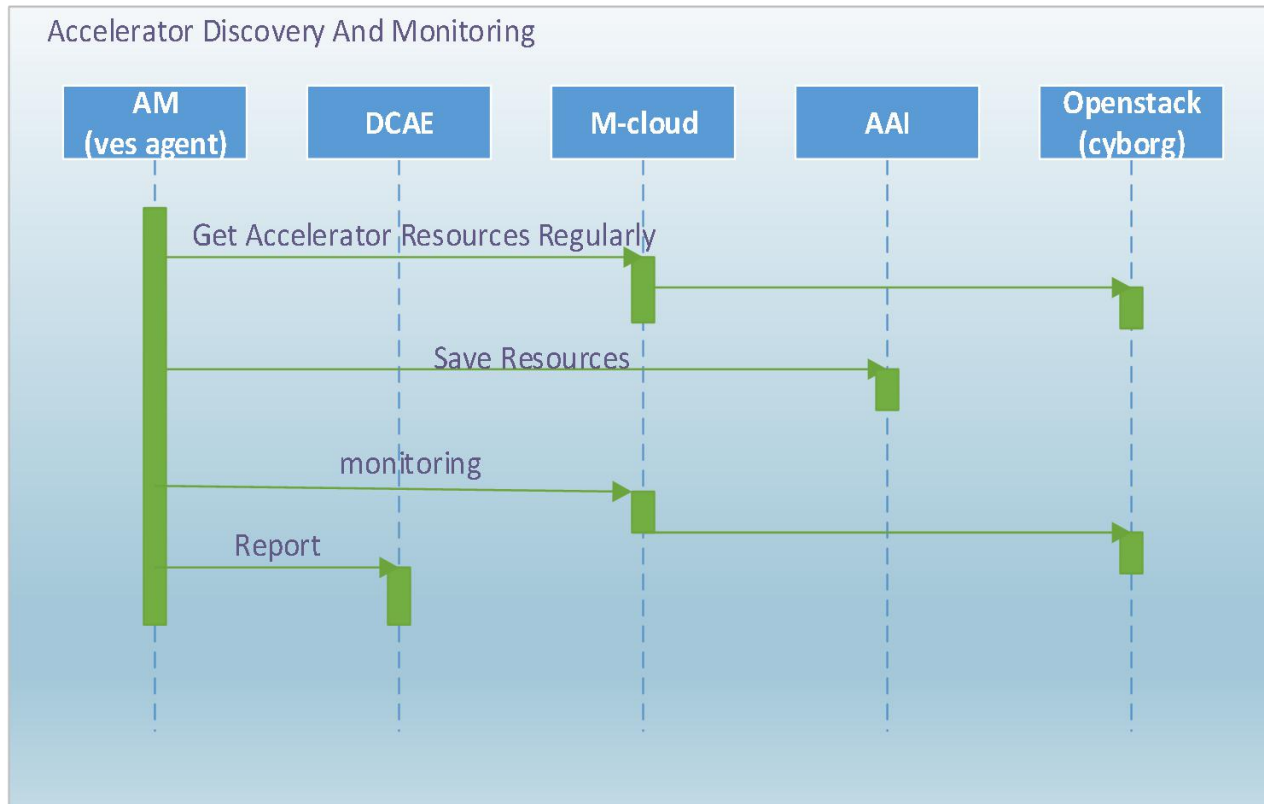
1. Discovery and registration of acceleration resources

2. Instantiation and assignment of acceleration resources (Reuse existing process in HPA)

3. The termination of the use of acceleration resources and release process.

4. Adjustment of suitable acceleration resources according to the data AM monitored and analyzed. – Whether to put it in D release or beyond depends on different target of different cases, seen as stretch goal.

1. Discovery and registration of acceleration resources

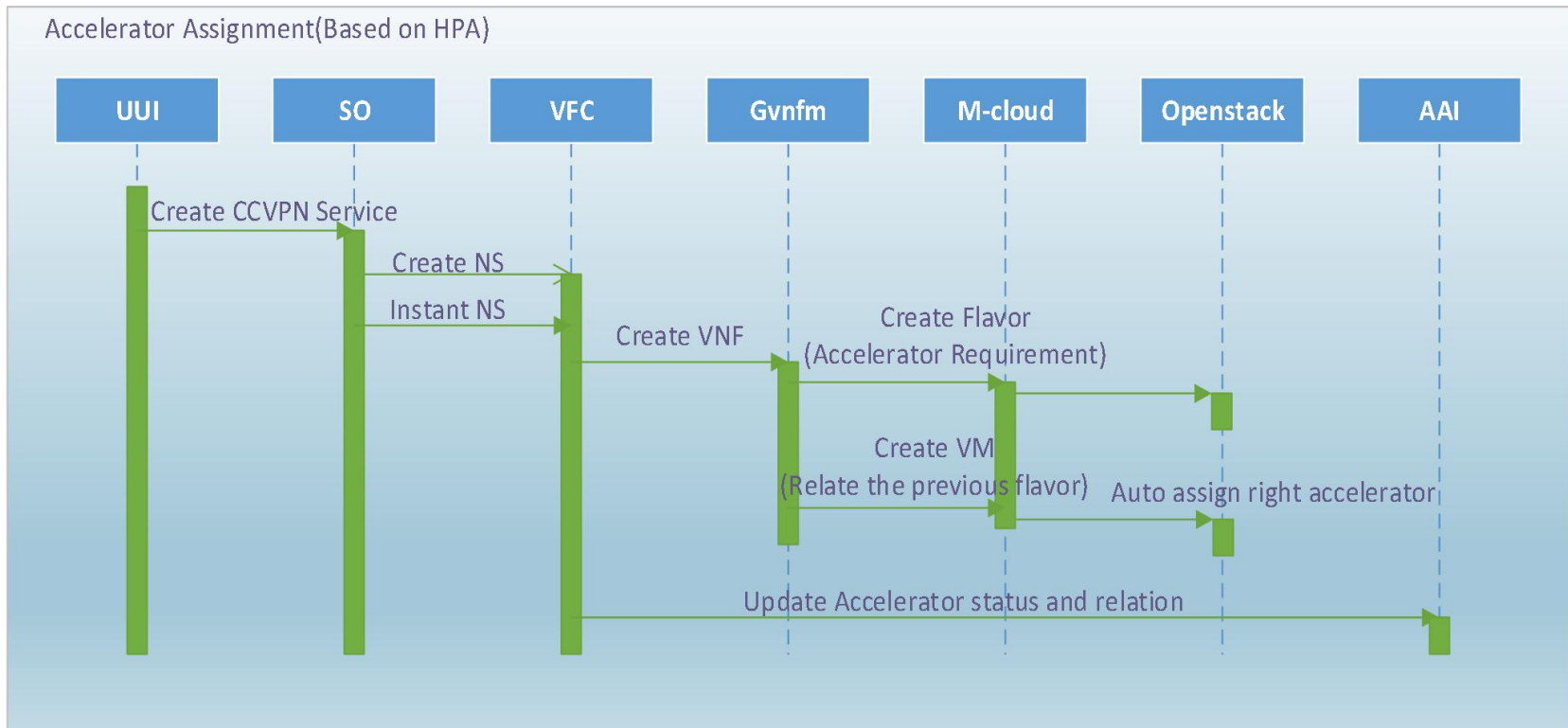


AM: Acceleration Management

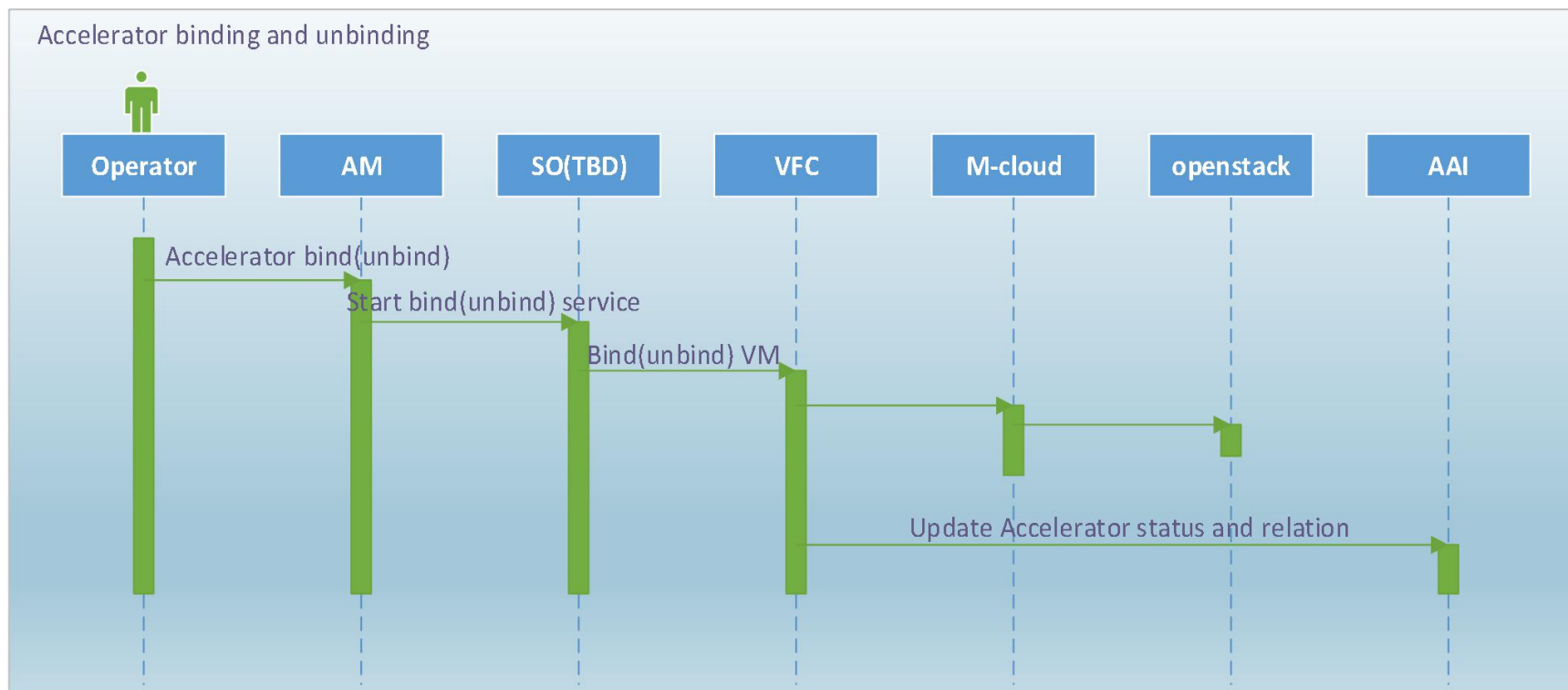
HPA upload all the accelerator informations when registering the vim to ESR. But if some accelerators become abnormal, AAI doesn't know it

2. Instantiation and assignment of acceleration resources

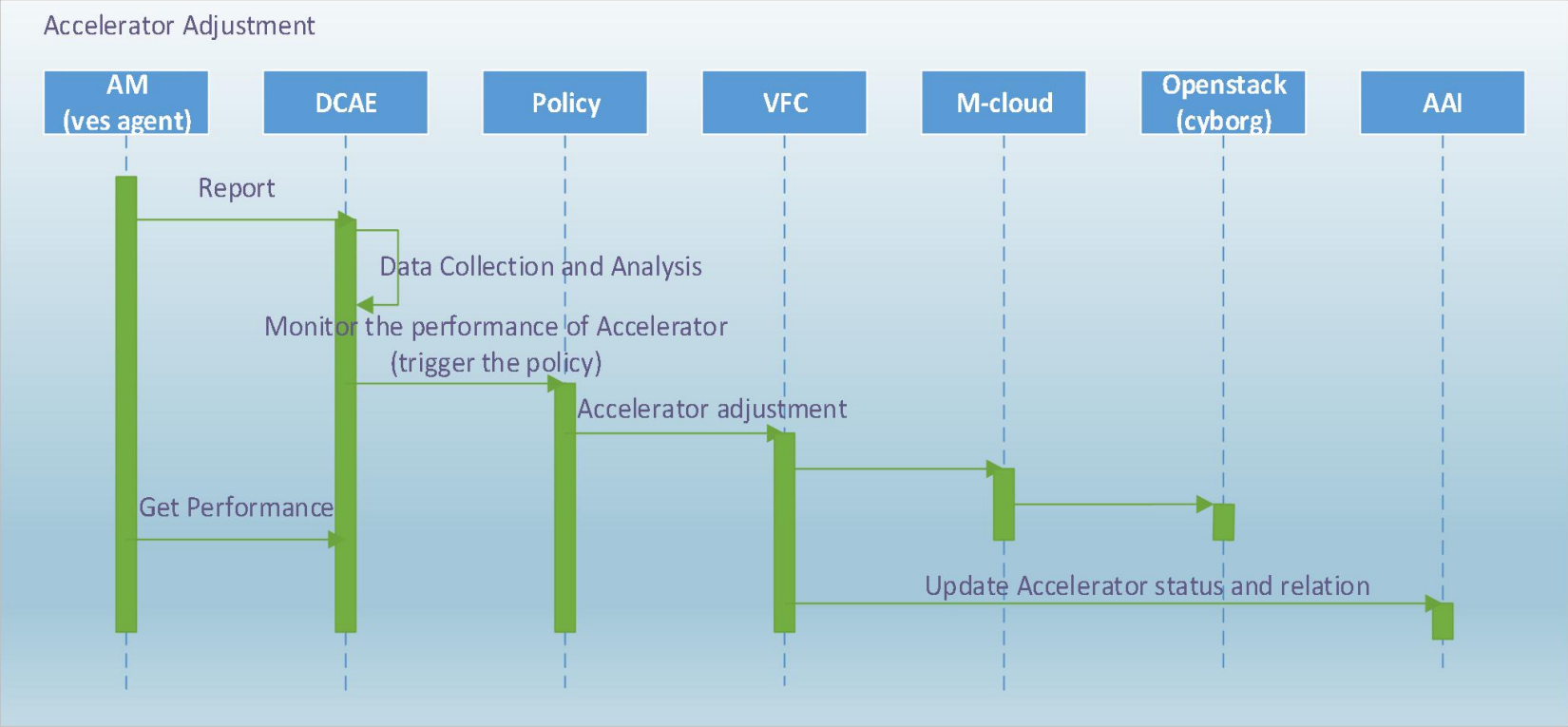
- Work with HPA and need its enhancement



3. Termination of the use of acceleration resources and release process



4. Adjustment of acceleration resources



BP of AM

- Divide it into many stages , do it step by step:
 1. Build GPU resource management capability in AM framework through the determined demo requirement in Barcelona mobile world congress, in this stage not following D release plan in ONAP community, and set it as offline stage .
 2. Through demo's practice, we have built AM framework. At this stage, we will introduce it into ONAP E release, incorporating more acceleration resource management capabilities into AM system, such as FPGA , etc. As well as support more cases .
 3. Indeed we have these potential cases:
 - 1) Offloading of HTTPS and TLS
 - 2) FPGA adjustable management to support AI cases
 - 3)

BP of AM

- About our regular meeting plan:
 1. Following demo progress every two weeks.
 2. Discuss the concerns around acceleration management from all of you, including HPA, Cyborg and so on.
 3. Discuss possible user scenarios based on the latest requirements.
 4. Discussion of the latest acceleration technologies.
 5. Cooperation with other opensource communities.

Demo preparation

Blue print, need more discussion:

- 1. Sharing demo implemented in Barcelona mobile world congress last week.
- 2. Discuss and determine the basic architecture of AM.
- 3. Build GPU resource management capability in AM framework.
- 4. Complete the test work.
- 5. Following demo progress every two weeks.

Call for volunteers

- Welcome to join in AM thread, send your email with 'Commit' or 'Contribute' in subject line.
- Contact info:
 - ✓ Name: Lei Huang
 - ✓ Email address: 18350830036@163.com
 - ✓ Wiki :
<https://wiki.onap.org/display/DW/Acceleration+Management+Interest+Group+Participants>