



Modèle de contrôle intelligent pour le Fog Cloud :

Vers une approche SDN intelligente pour le traitement des données massives au service de la radioastronomie

Imène ELLOUMI ZITOUNA
Associate Prof. at ENIT
09/02/2021

Tatooine planet World, le laboratoire Sys-com, la cité des sciences, le laboratoire d'études spatiales et d'instrumentation en astrophysique, la STU-paris et le Projet Radiotélescope en Tunisie organisent

Journées de Sensibilisation franco-tunisienne sur

la radioastronomie

8-9 février 2021
8h30-17h30
Cité des sciences de Tunis

Présentation en visioconférence :
- Des aspects scientifiques et techniques de la Radioastronomie, Des compétences et des expériences des laboratoires Tunisiens et Français

INTERVENANTS par ordre alphabétique

| | | |
|------------------|-----------------|-----------------|
| Toufik Aguel | Silviane Corbel | Azza Ouled Zaid |
| Yahya Ben Salah | Monica Elcorani | Philippe Sakoné |
| Faouzi Boussaha | Chiara Ferrari | Jeanne Treutzel |
| Kamel Charbib | Julien Girard | Gilles Theureau |
| Didier Charrier | Moncef Ghourabi | Cédric Viou |
| René Cogard | Meriem Kassar | Martina Wiedner |
| Françoise Combes | Fethi Meiri | Philippe Zarka |

PROGRAMME : radio-tun.tun.cienciaeconf.org

partenaires : RADIOTELESCOPE EN TUNISIE, Observatoire de Paris - LESIA, PSL Université Paris, Radiotélescope de France, STU, ARAB COM, TATOOINE, Institut National de l'Électronique et de l'Informatique de Tunisie

Résumé

Avec les big data et le cloud computing, les bases d'échantillons, le problème de classification, d'enregistrement et de monitoring de signaux naturels en présence d'interférences intenses généré dans les observations radioastronomiques serait extraînement favorable à un traitement de networking intelligent basé sur de la qualité de service au service d'une analyse de données prédictive.

Les big data sont des données issues d'un flux très grand en termes de volume, très rapides (vélocité de la donnée). Elles proviennent de différentes sources et sous différentes formes (Variété de la donnée). Cette combinaison de facteurs rend extrêmement difficile la capacité d'analyser et traiter la donnée en direct. Néanmoins, une telle capacité à faire « parler » la donnée immédiatement est exactement ce qui est recherché aussi dans un domaine tel que la Radioastronomie. Avec la virtualisation du réseau défini par logiciel SDN son architecture en couches et ses interfaces programmables, le contrôle global du réseau de transport de ces données est par conséquent assuré par la centralisation logique de la fonction de contrôle au niveau des différents contrôleurs SDN. Les flux de réseau sont ainsi contrôlés à un niveau d'abstraction global à l'aide de protocoles comme par exemple l'OpenFlow.

Nous présentons, à travers cette contribution, un nouveau modèle d'apprentissage Intelligent et hybride à la fois supervisé et par renforcement dans une infrastructure de softwarisation SDN. Qui permet d'ajouter de l'apprentissage à la fonction de networking et de contrôle pour une meilleure prise de décision intelligente au niveau des contrôleurs SDN afin de router de la façon la plus intelligente ces données sur les réseaux. Puis récupérer les flux de données et transmettre ces fonctionnalités à un neuronal convolutif modèle de réseau pour séparer les classes. La tâche consiste à extraire les informations pertinentes du flux de données générés à partir d'un réseau défini par les logiciels et de les utiliser pour identifier le nœud de destination le plus optimal avant de transmettre aux serveurs big data. Le résultat est prometteur, le modèle a atteint une précision de 72,7% sur une base de données de 16 classes

Mots clés : Big data, Cloud Computing, SDN, OpenFlow, Intelligence artificielle, Reinforcement learning, Virtualisation.

Plan

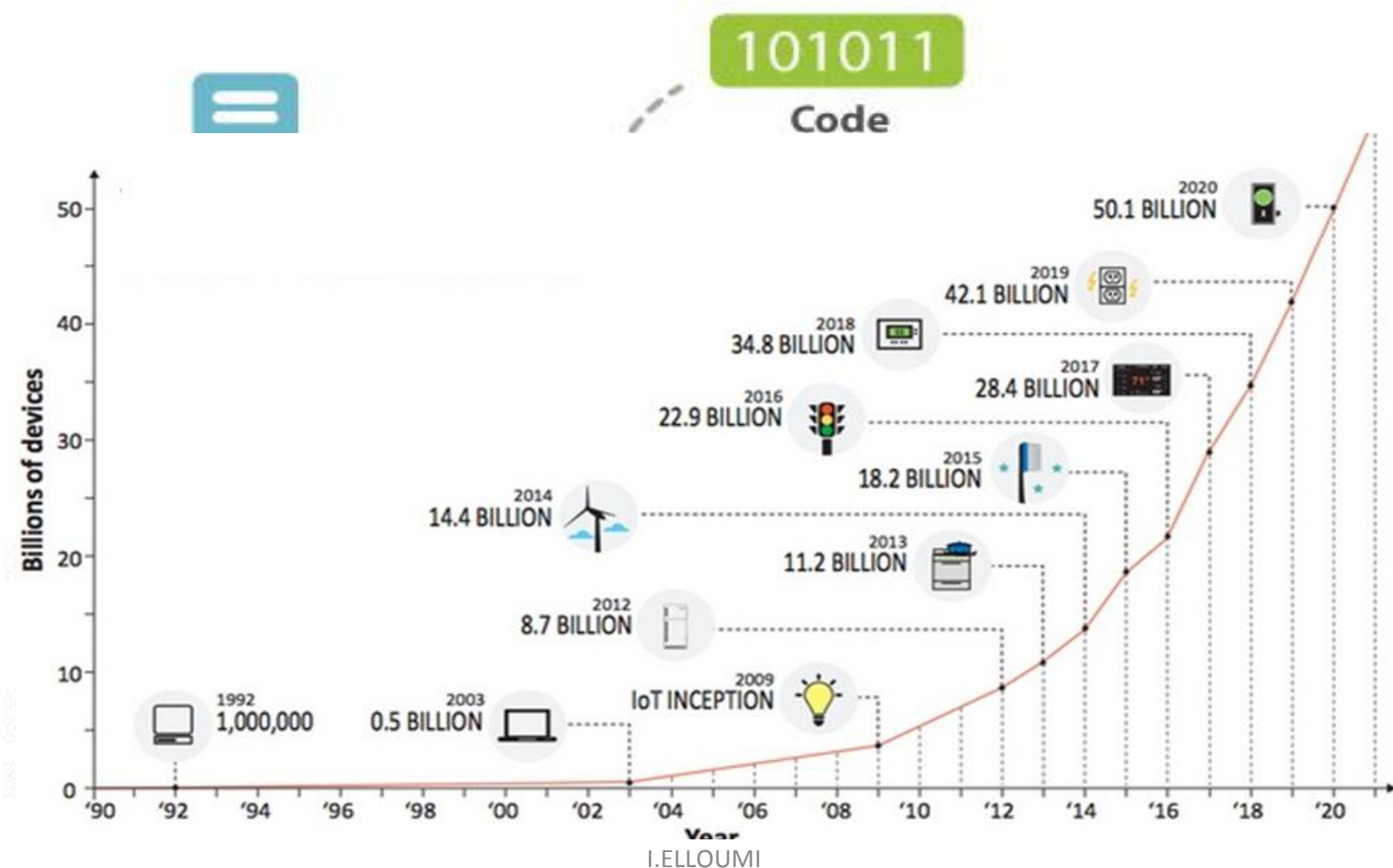
- I. Objectifs & Introduction générale
- II. Virtualisation
- III. Virtualisation du réseau
- IV. Problématique
- V. Software Defined Networking - SDN
- VI. Machine Learning
- VII. Approche SDN intelligente
- VIII. Vers une approche SDN intelligente pour le traitement des données en radioastronomie
- IX. Conclusion générale

Objectifs

1. Comprendre les concepts de réseautique définie par logiciel (SDN) ainsi que les motivations et les avantages offerts par cette technologie pour gérer les données volumineuses issues de différentes sources.
2. Connaître les différentes applications possibles de la technologie SDN pour traiter les données prioritaires avec différents niveaux de QoS.
3. Présenter la contribution qui consiste à intégrer de l'intelligence aux contrôleurs SDN grâce à l'application des algorithmes d'apprentissage automatique en terme d'optimisation du routage, classification du trafic et gestion des ressources prioritaires à la demande.

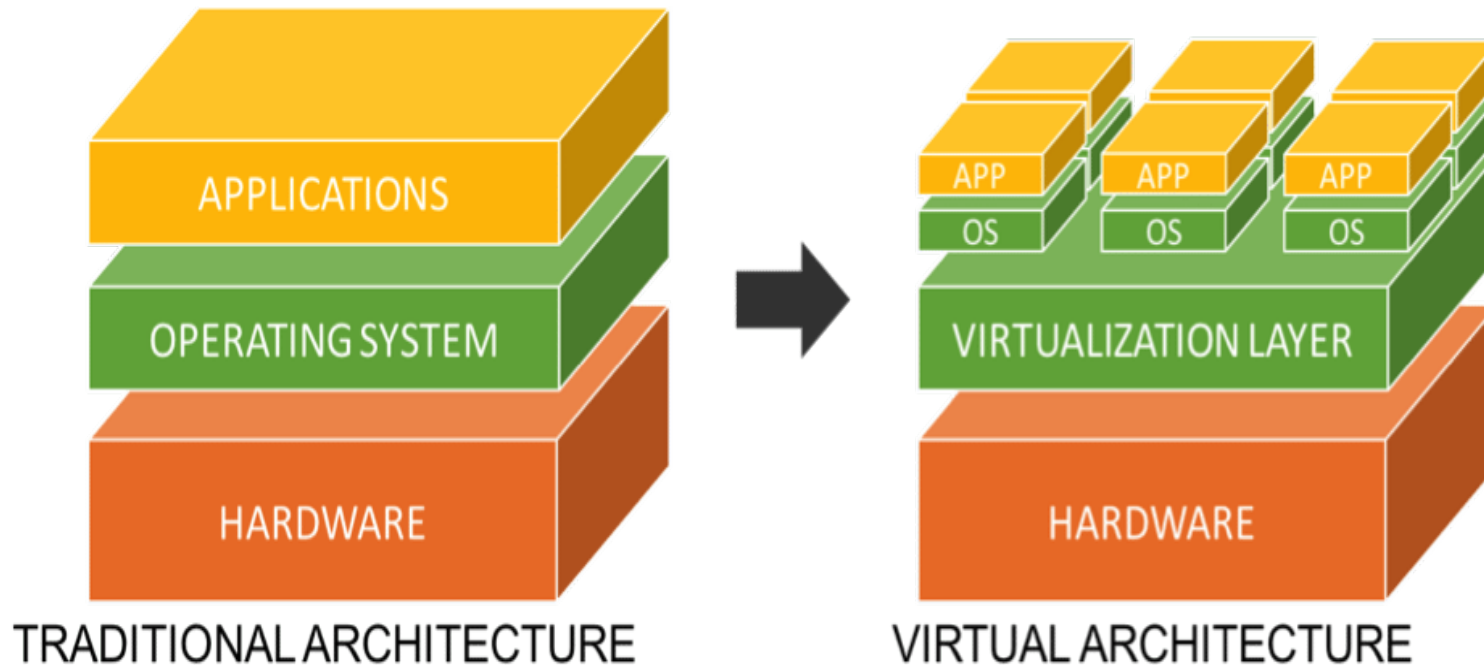
Introduction générale

- ❖ Croissance des multiples applications dynamiques tirées par l'essor des outils populaires tel que le **Cloud Computing**, **Fog Computing** et l'**Internet des Objets**.



Virtualisation

1. Concept globale de la virtualisation



2. Types de virtualisations: **Virtualisation complète** ; Para-virtualisation ; Virtualisation du réseau

Virtualisation

Réseaux traditionnels Fonctionnement d'un routeur

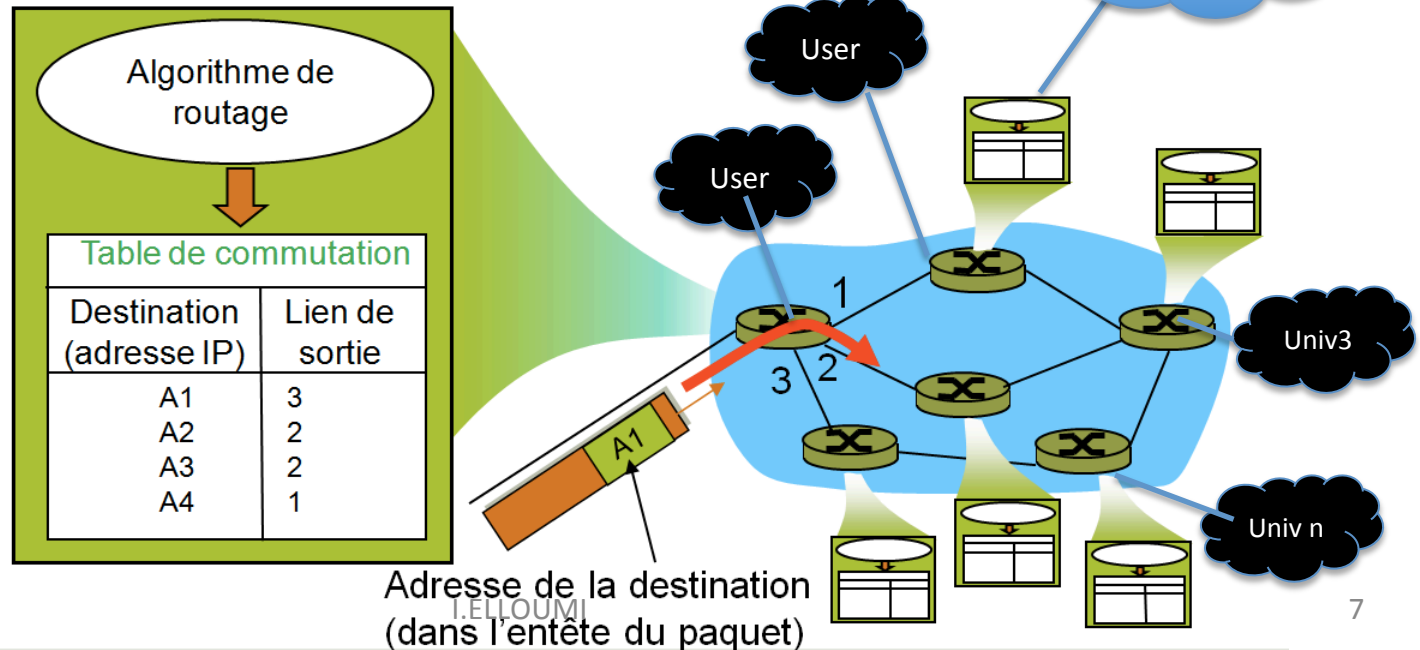
- Routage : déterminer l'itinéraire (route) qui doit être empruntée par les paquets.
 - Algorithms de routage distribués et "autonomes" comme OSPF, RIP
- Commutation (forwarding): déplacer les paquets d'une entrée du routeur à la sortie du routeur appropriée

-Très difficile de définir des règles de commutation ou de routage personnalisées

-Impossible de définir une QoS à une granularité très fine

-Technologies propriétaires (Vendor Lock-in) et fermées

-Automatisation difficile



Virtualisation

Propriétés des réseaux virtuels

- Partitionnement: chaque ressource peut être utilisée par plusieurs réseaux virtuels.
- Isolation: les réseaux virtuels sont isolés les uns des autres.
- Abstraction: une ressource virtuelle ne correspond nécessairement pas à la même ressource physique.
- Agrégation: plusieurs instances de réseaux virtuels peuvent être agrégées pour créer un réseau plus grand (plus de ressources).

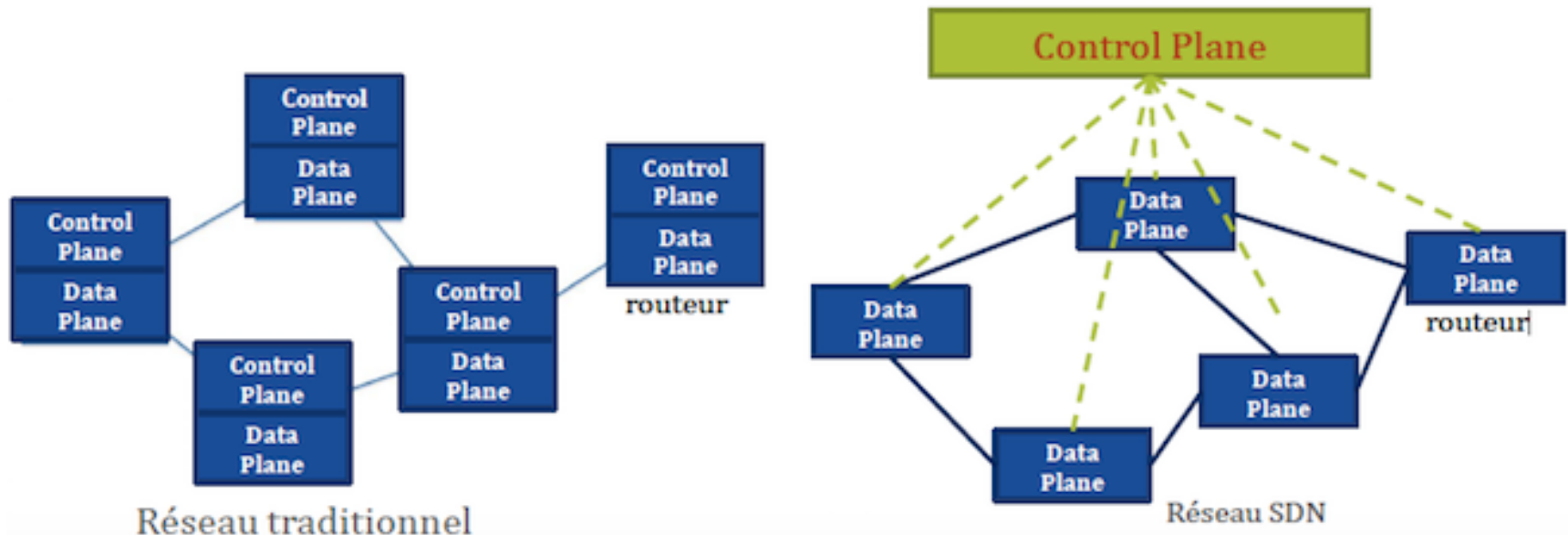
Virtualisation

Avantages des réseaux virtuels

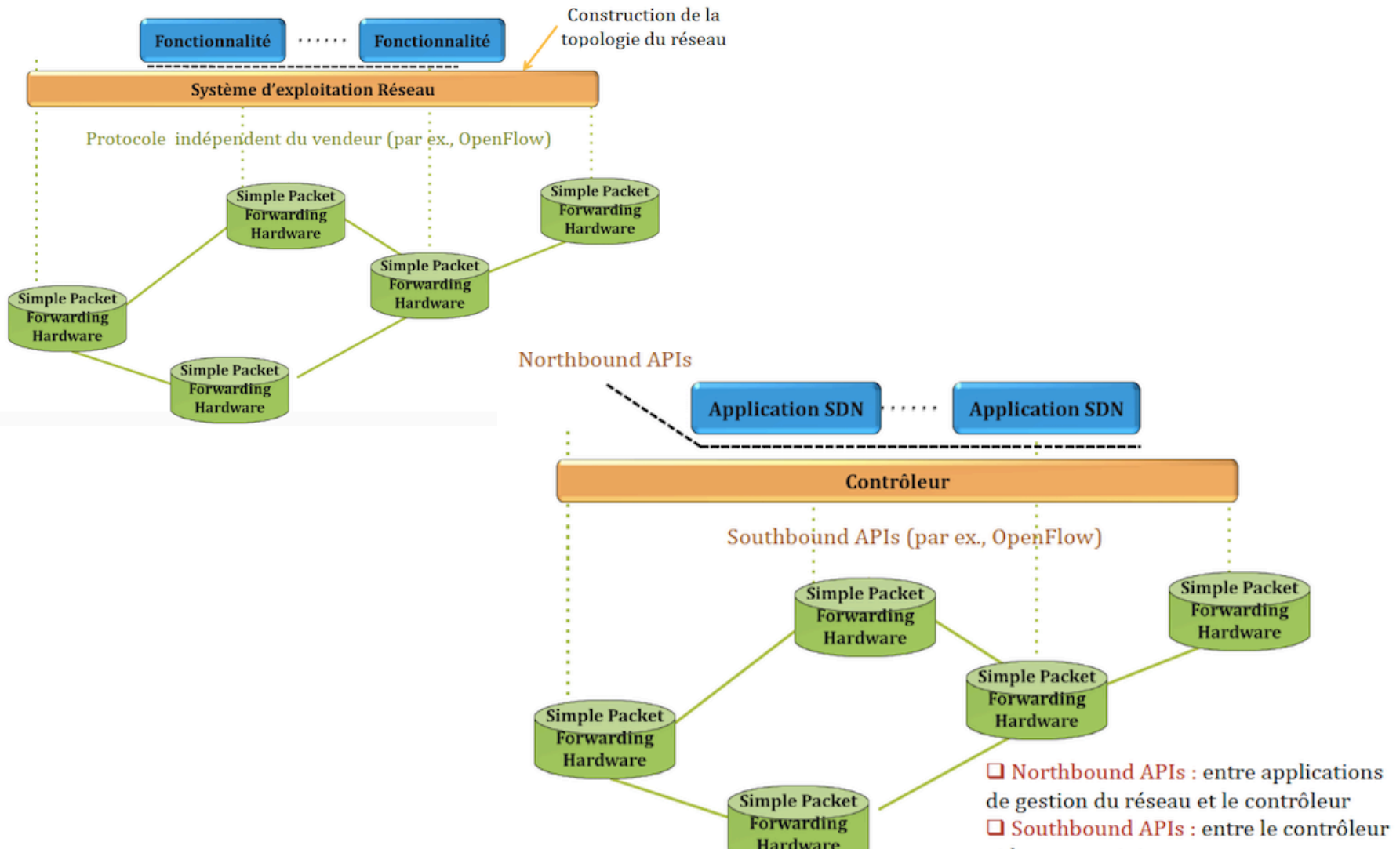
- Chaque réseau virtuel peut utiliser sa propre technologie, algorithme de routage, mécanisme de gestion de flux indépendamment des réseaux virtuels.
- On peut facilement ajouter des ressources, modifier la topologie et la configuration d'un réseau virtuel à la demande ou de façon dynamique.

Software Defined Networking (SDN)

- Nouvelle technologie permettant de configurer d'une façon dynamique et rapide les équipement réseaux
- Toute l'intelligence est déplacée à un contrôleur central



Software Defined Networking (SDN)



Software Defined Networking (SDN)

-Plan décisionnel est séparé du plan opérationnel

→ Les routeurs et les commutateurs deviennent simplement des équipement de commutation simples (appelés commutateurs SDN)

→ Le plan décisionnel est centralisé dans un contrôleur

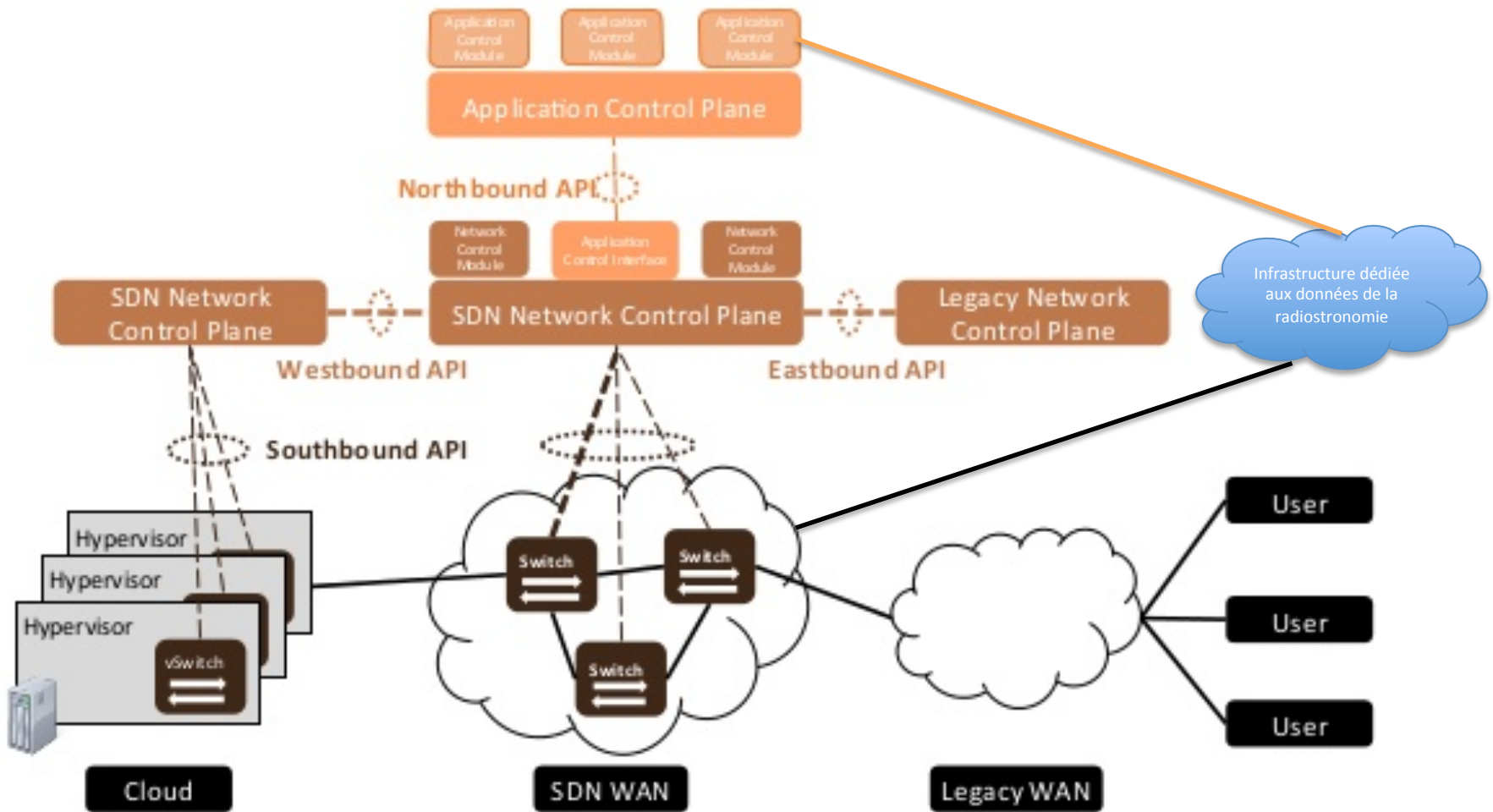
-Le contrôleur a une vue globale du réseau & prend des décisions de commutation et programme les commutateurs

-Le plan décisionnel peut être implémenté dans un seul contrôleur ou un ensemble de contrôleur qui coordonnent entre eux.

→ Des interfaces ouvertes sont définies entre le contrôleur et les commutateurs.

Software Defined Networking (SDN)

Interfaces et protocoles d'une architecture SDN :



Software Defined Networking (SDN)

-Gestion du trafic par flux:

- Flux (flow) : ensemble de paquets partageant un ensemble de caractéristiques
- Avec SDN, chaque flux est défini et routé d'une façon indépendante des autres

→ Permet de définir **des règles de routage et de commutation à une granularité très fine**

| Switch Port | MAC src | MAC dst | Eth type | VLAN ID | IP Src | IP Dst | IP Prot | TCP sport | TCP dport | Action |
|-------------|---------|----------|----------|---------|--------|--------|---------|-----------|-----------|--------|
| * | * | 00:1f:.. | * | * | * | * | * | * | * | port6 |

Software Defined Networking (SDN)

Avantages du SDN pour orchestrer les données volumineuses :

-Pas de dépendance au vendeur (Vendor Lock-in)

- *Le même contrôleur peut gérer n'importe quel équipement
- *Les applications de gestion de trafic sont indépendantes des switches

-Open-Source

- *Facilité de développement et déploiement de nouvelles applications pour la gestion des réseaux
- *Facilité de mettre à jour le logiciel gérant le réseau

-Gestion centralisée: il est plus facile d'optimiser le routage du trafic dans le réseau grâce à la vue globale disponible au niveau du contrôleur

- *Routage Ingénierie de trafic (par ex., équilibrage de charge)
- *Meilleure décision en cas de pannes (d'un lien ou d'un nœud) & Equilibrage de charge

-Flexibilité, automatisation et programmabilité

- *Facilité de reconfigurer les commutateurs
- *Possibilité de définir une QoS à une granularité très fine

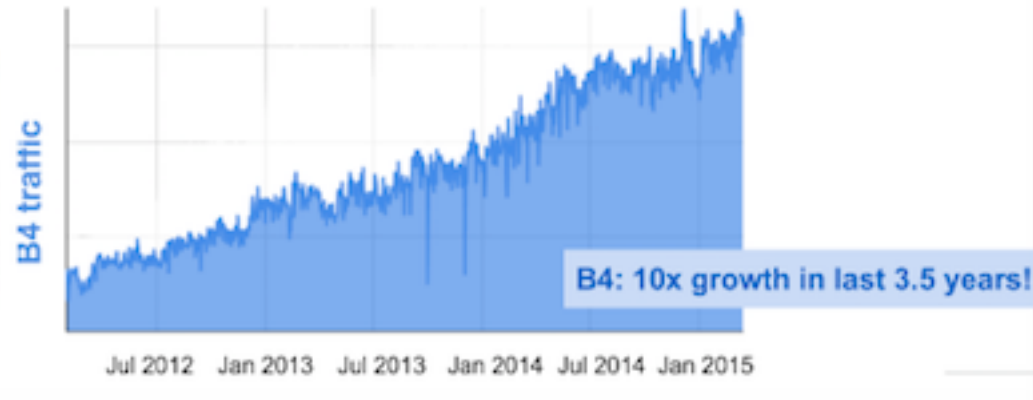
-Virtualisation du réseau : diviser le réseau en plusieurs réseaux virtuels

Software Defined Networking (SDN)

Exemple d'application du SDN : Le réseau Google B4

- B4 assure la connectivité entre les centres de données de Google (2011)
B4 transporte principalement des réplicas de données (photos, vidéos, ..)

→ Volumes de données importants



B4 utilise la technologie SDN avec le protocole OpenFlow

Ingénierie de trafic centralisée

- *Maximiser l'utilisation des liens tout en évitant leurs congestions
- *Routage Multipath et réservation de route en fonction de la priorité de l'application
- *Allocation dynamique de la bande passante

Plusieurs liens du réseau B4 opèrent à presque 100% d'utilisation pour longues durées

Virtualisation du Networking

Problématique

❖ **Absence de mécanisme d'évaluation et de contrôle de QoS** : choix du **meilleur contrôleur SDN et meilleur Vswitch** en terme de temps de réponse dans des situations de congestion du réseau et aussi la centralisation au niveau du contrôleur opérée par le protocole OpenFlow permet d'augmenter ces temps de réponses avec des risques de DoS même pour les solutions virtuelles.

❖ **La surcharge de flux de données au niveau du contrôleur peut rendre l'opération de contrôle, la gestion ainsi que la classification du trafic inefficace** : des temps de traitement d'autant plus importants pour des flux volumineux de données.

⇒ **Besoin d'apporter certaine intelligibilité au contrôleur**: sans **mécanisme de prédiction qui soit intelligent**, il est simple pour un contrôleur adverse d'usurper le rôle de maître, ou d'empêcher qu'un contrôleur légitime qui souhaite être maître puisse paramétrer le Vswitch, ou du moins avoir des temps de traitement qui augmente avec le taux de trafic.

⇒ **Recours à une des techniques de l'intelligence artificielle hybride, celle de l'apprentissage par renforcement et l'apprentissage supervisé afin de résoudre ce problème.**

IA & Machine Learning

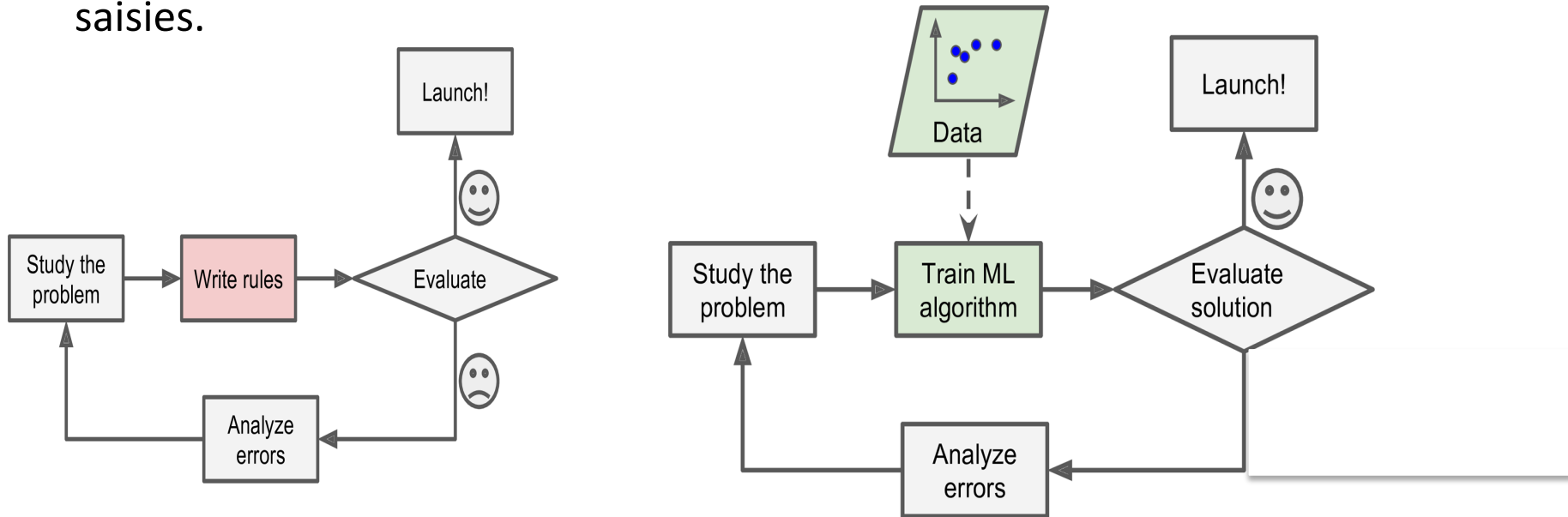
1. Machine Learning
2. Importance de la machine learning
3. Approche traditionnelle vs approche de la machine learning
4. Différents catégories de la machine learning

Machine Learning

- La Machine Learning peut être défini comme étant une technologie d'intelligence artificielle permettant aux machines d'apprendre sans avoir été au préalable programmées spécifiquement à cet effet.
- Le Machine Learning est explicitement lié au Big Data, étant donné que pour apprendre et se développer, les ordinateurs ont besoin de flux de données à analyser, sur lesquelles s'entraîner.
- Le Machine Learning est actuellement très utilisé pour le Big data dans différents domaines: diagnostic médical, remplacement d'images dans les vidéos, domaine de la sécurité, etc.

Approche traditionnelle vs approche de la machine Learning

- ❖ L'apprentissage automatique simplifie l'utilisation des ordinateurs dans la construction de modèles à partir de données d'échantillonnage en vue d'automatiser les processus de prise de décision en fonction des données saisies.



Apprentissage supervisé

Apprentissage non supervisé

L'apprentissage semi-supervisé

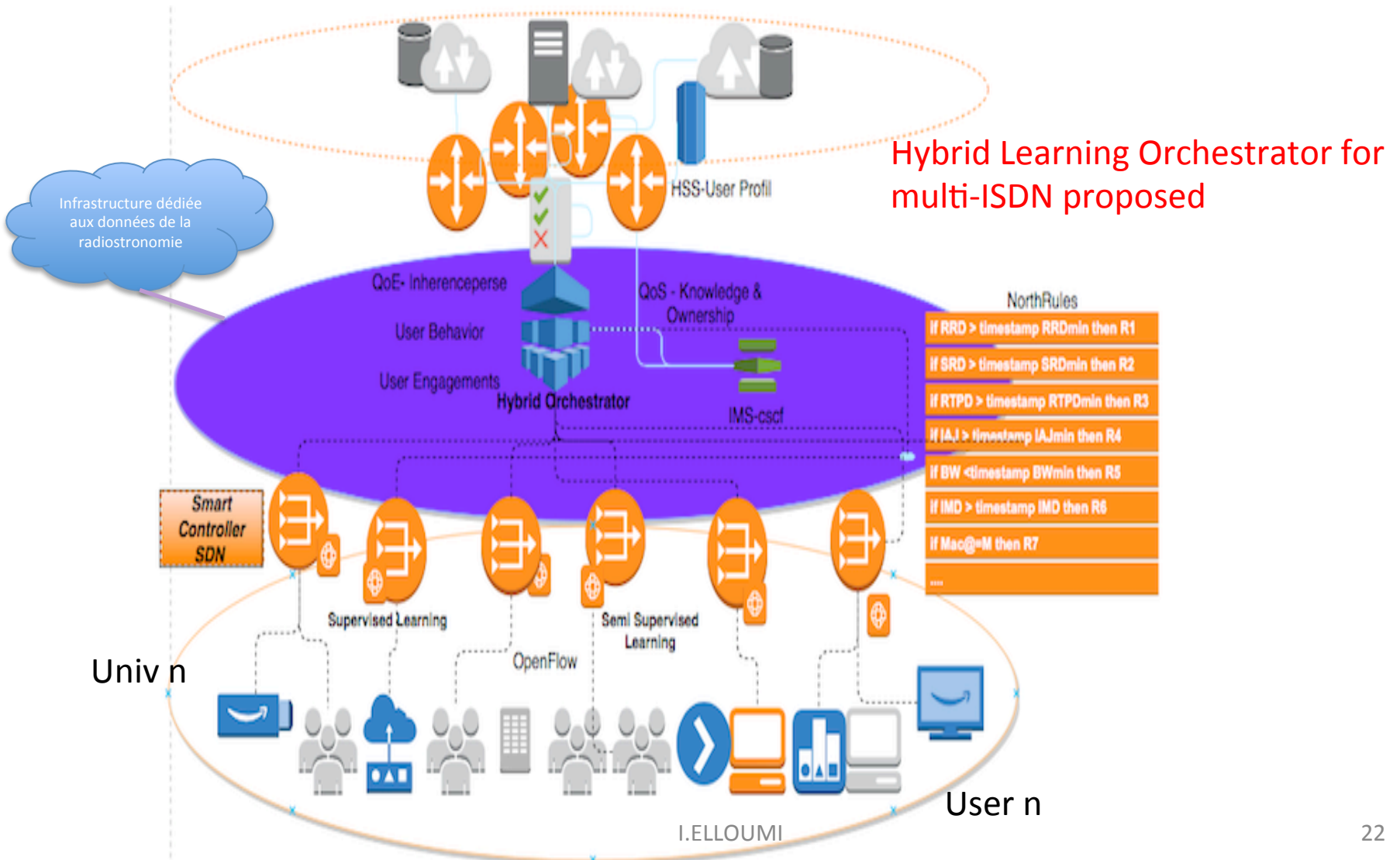
Apprentissage par renforcement

Approche SDN intelligente

1. Intelligence artificielle apportée aux traitements des données dans SDN avec des niveaux de QoS.
2. Choix de l'apprentissage par renforcement et de l'apprentissage supervisé
3. Utilisation de l'algorithme Q-Learning et des Réseaux de neurones
4. Elaborer une politique de maximisation des récompenses cumulatives à long terme obtenues via des transitions d'état en se basant sur les actions, les états et les Q-values
5. Base de données (Dataset)

Approche SDN intelligente proposée

Approche ISDN



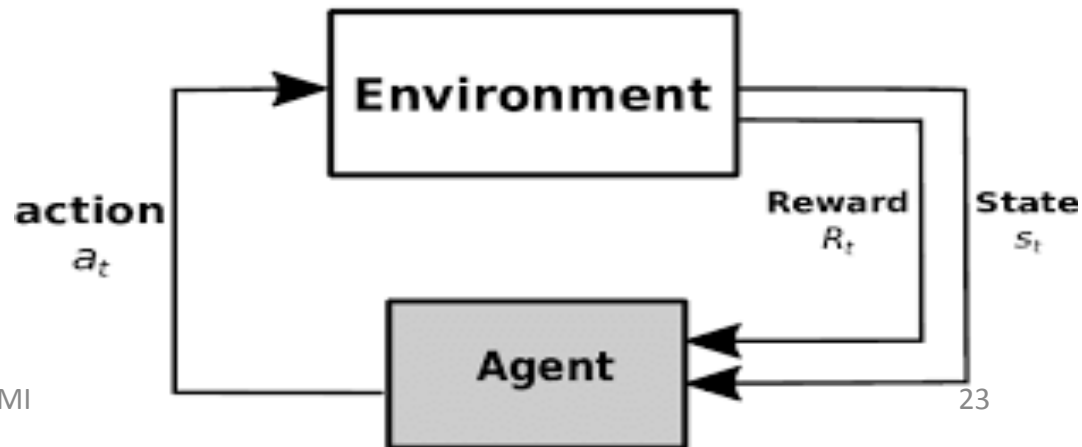
Approche SDN intelligente proposée

Approche ISDN

Modélisation d'apprentissage par renforcement

Enchaînement :

- ❖ Création de l'environnement
- ❖ Création de la Q-table et initialisation des Q-values (on assimile des résultats des actions effectuées par l'agent par des valeurs notés Q-values)
- ❖ Application de l'algorithme de Q-learning
 - 1/ Choix de l'action et équilibre exploration vs exploitation
 - 2/ Application de l'action dans l'environnement et récompense
 - 3/ Mise à jour des Q-values en ayant recours aux équations de Bellman, puis stocker ces valeurs dans un tableau noté Q-table



Approche SDN intelligente proposée

Approche ISDN

Base de données (Dataset)

- ❖ Prévoir lequel, parmi plusieurs commutateurs, est plus performant en terme de QoS, tout en jouant sur les différents valeur de TCAM de ces commutateurs, dont le pourcentage de valeur est compris entre 0 et 90.

- ❖ Cet intervalle de valeur, peut être divisé en quatre états différents comme suit :
 - Etat 1 : Si la valeur de la TCAM appartient à $[0..30]$: il s'agit d'un état faible
 - Etat 2 : Si la valeur de la TCAM appartient à $[30..50]$: il s'agit d'un état apprécié
 - Etat 3 : Si la valeur de la TCAM appartient à $[50..80]$: il s'agit d'un état critique
 - Etat 4 : Si la valeur de la TCAM appartient à $[80..90]$: il s'agit d'un état médiocre

Approche SDN intelligente proposée

Approche ISDN

Utilisation de l'algorithme Q-Learning

La lettre Q fait appel à la mesure de la qualité d'une action effectuée dans un état donné par le système à prédire.

En partant d'un état initialisé du système, la fonction de valeur état-action notée Q permet de déterminer la récompense sur le long terme $Q(s,a)$.

- Equation de Bellman modifiée :

$$Q(s_t, a_t) = r + Q(s_t, a_t) - \epsilon Q(s_t, a_t) + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1})$$

- Mise à jour de chaque valeur obtenue de Q dans un tableau appelé Q-Table renfermant les états du système et les différentes actions qui peuvent subir.
- Le taux d'apprentissage γ , fixé au fur et à mesure, qui détermine dans quelle mesure les informations nouvellement obtenues l'emportent sur les anciennes.
- Comme l'apprentissage par renforcement est itératif, on répète les opérations ainsi citées jusqu'à obtenir un résultat souhaité.

Approche SDN intelligente proposée

Approche ISDN

Algorithme de Q-learning proposé

ALGORITHM : Algorithme de Q-Learning proposé

Preconditions:

Initialize Q-table random TCAM percentage

Initialize state s_t

Initialize goal ϵ

Procedure:

01: **repeat**

02: **for** (step = 0; step < learning_iteration; step++)

03: Get action a_t from s_t using β

04: Get parameter from s_t using β

05: $\beta = \beta - \beta * (\text{step} / \text{learning_iteration})$

06: Take action a_t on the parameter and receive reward r ,
control overhead C

07: Sample new state s_{t+1} after applied action a_t

08: Update $Q_t \leftarrow r + Q_t - \epsilon Q_t + \gamma * \max Q_{t+1}$

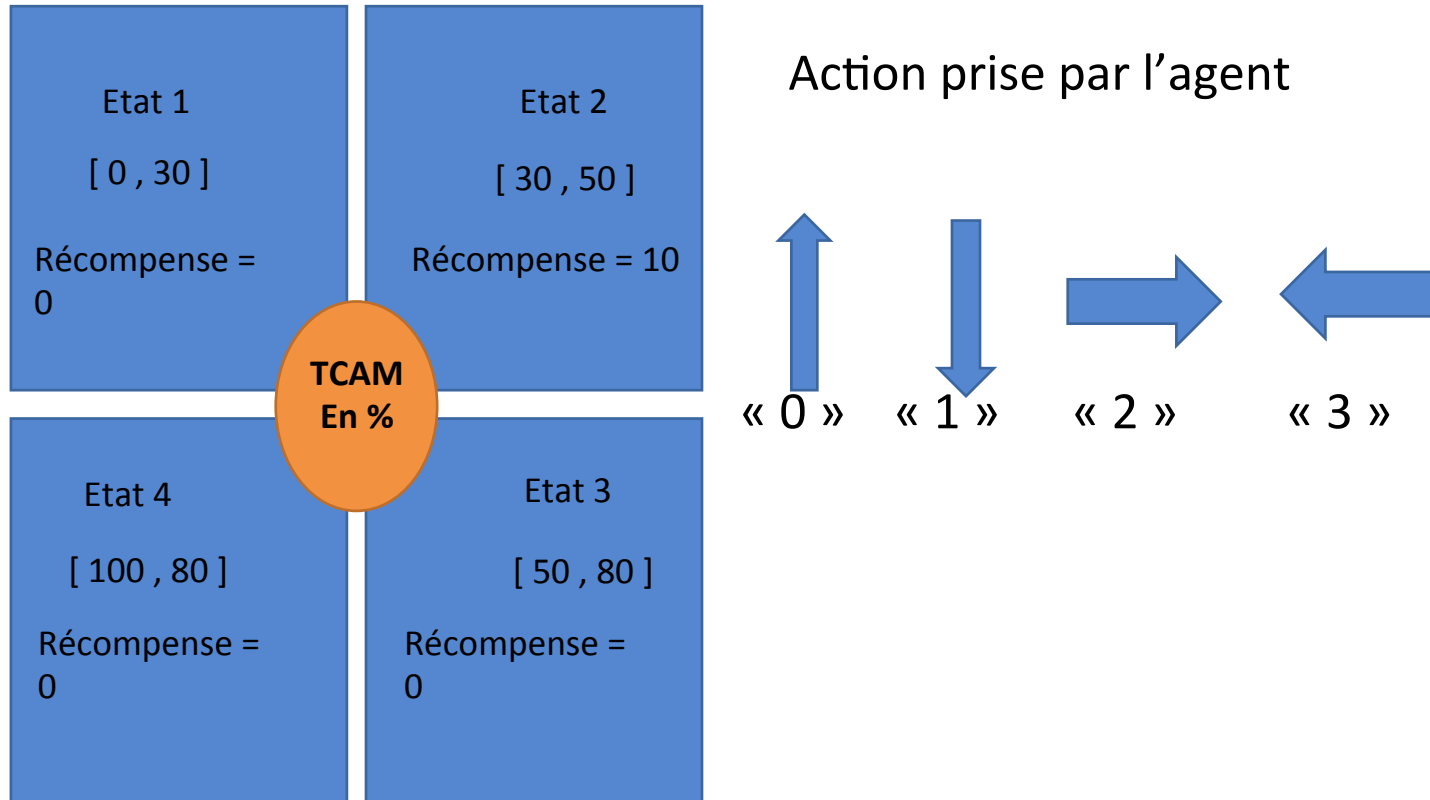
9: Update the corresponding parameter of Q_t

10: **end for**

Approche SDN intelligente proposée

Approche ISDN

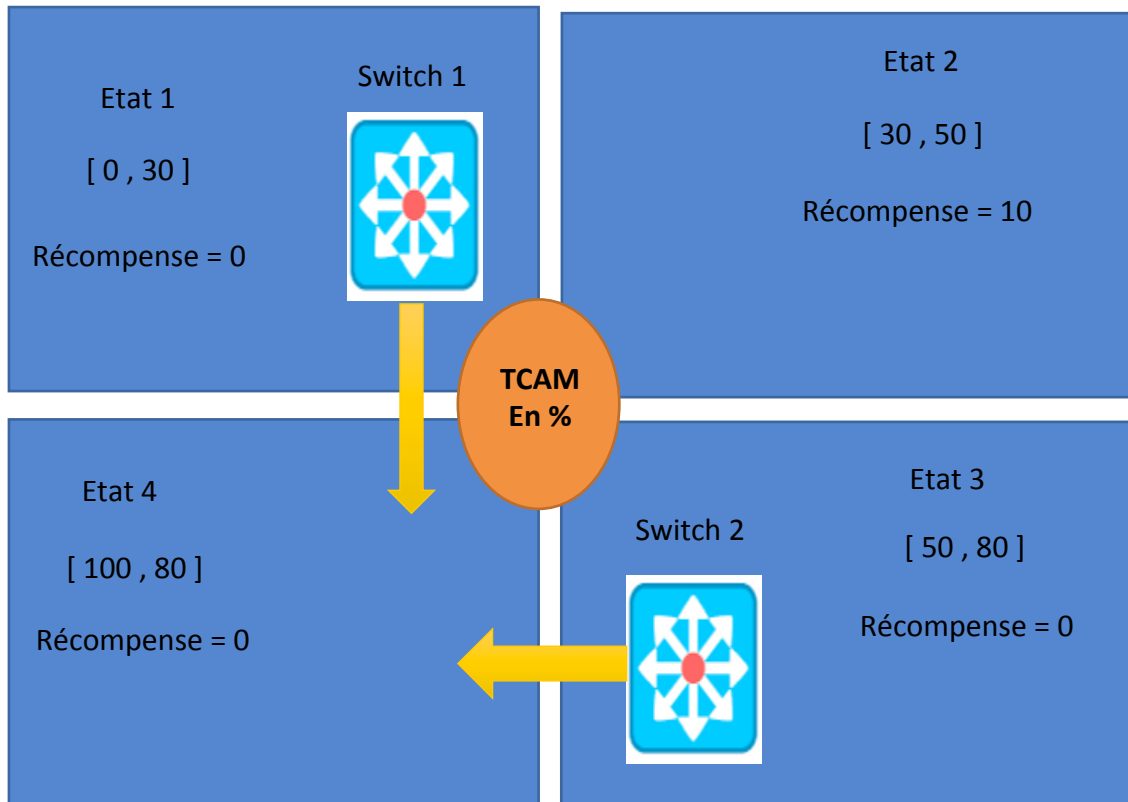
Environnement de l'agent



Approche SDN intelligente proposée

Approche ISDN

Résultats

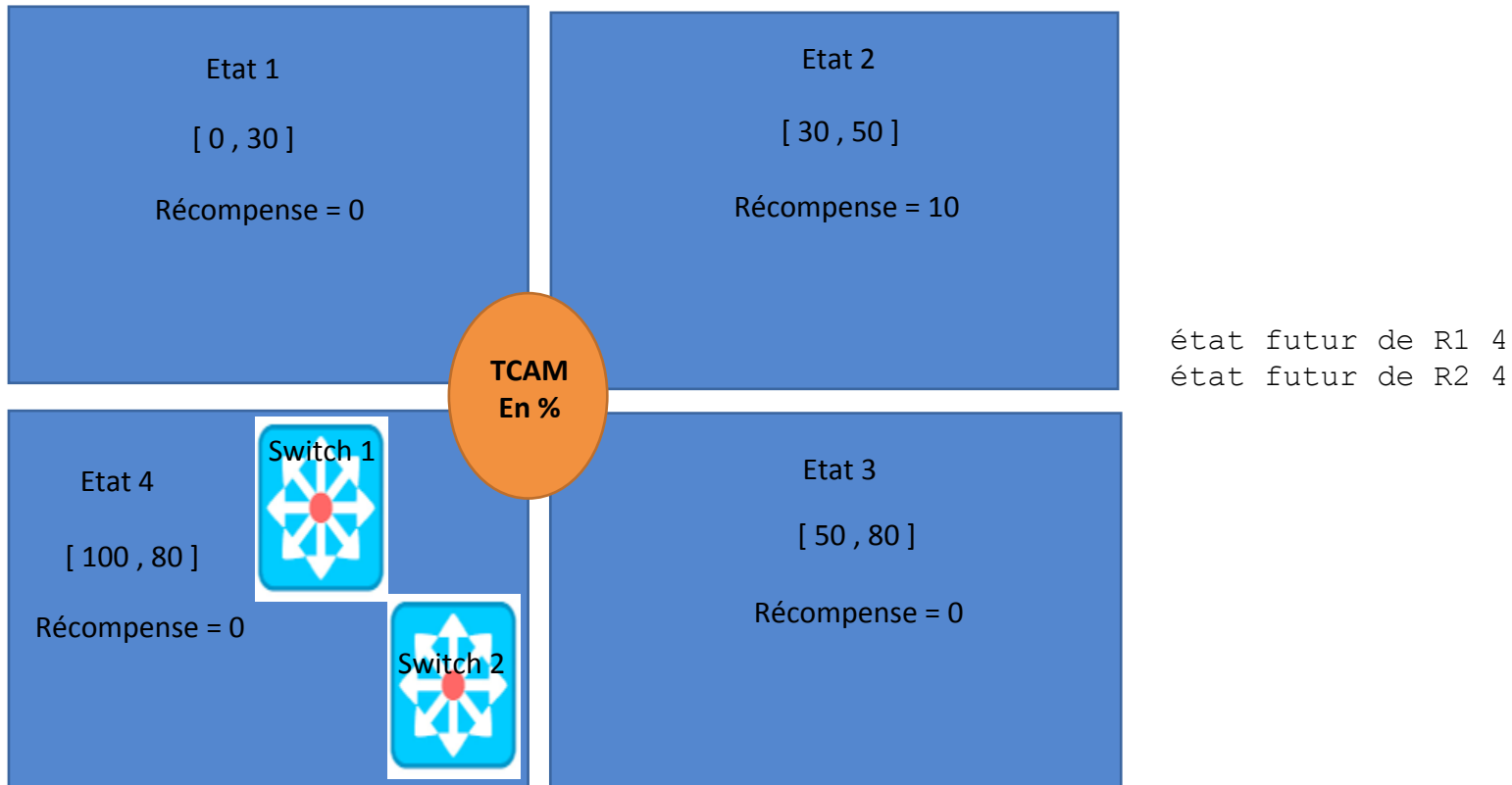


état de R1 1
état de R2 3
action sur R1 :
1 (DOWN)
action sur R2 : 3
(LEFT)

Approche SDN intelligente proposée

Approche ISDN

Résultats



Pour chaque action, le commutateur reçoit une récompense.
Mise à jour de Q-table.
Itération suivante.

Approche SDN intelligente proposée

Approche ISDN

Résultats

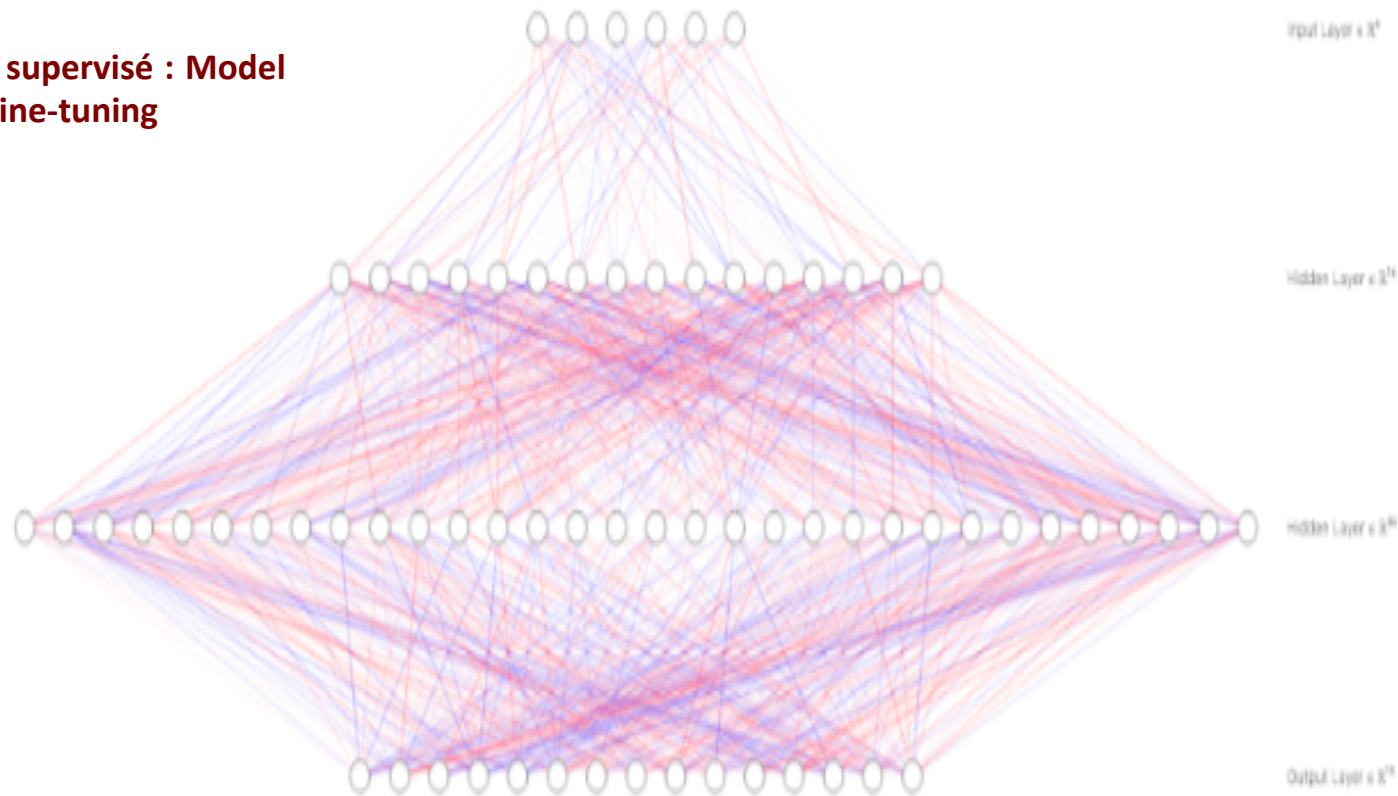
→ l'agent prédit le commutateur le plus fiable en terme de mémoire TCAM pour pouvoir gérer le trafic de données prioritaire : (R2)

```
etat de R1 1
etat de R2 3
action sur R1: 1
action sur R2: 3
etat futur de R1 4
Q_table de ROUTER 1 [[0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0]]
etat futur de R2 4
Q_table de ROUTER 2 [[0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0]]
etat de R1 4
etat de R2 4
action sur R1: 2
action sur R2: 0
etat futur de R1 3
Q_table de ROUTER 1 [[0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0]]
etat futur de R2 1
Q_table de ROUTER 2 [[0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0]]
etat de R1 3
etat de R2 1
action sur R1: 3
action sur R2: 2
etat futur de R1 4
Q_table de ROUTER 1 [[0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0]]
etat futur de R2 2
Q_table de ROUTER 2 [[0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 10.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0], [0.0, 0.0, 0.0, 0.0]]
R2 est le commutateur choisi
```

Approche SDN intelligente proposée

Approche ISDN

Apprentissage supervisé : Model Creation and Fine-tuning



The Input Layer: cette couche contient nos valeurs d'entrée avec lesquelles nous voulons prédire et entraîner le modèle, où chaque neurone de cette couche contient une valeur extraite de l'échantillon. Dans notre cas, notre jeu de données contient 6 colonnes, donc les 6 neurones de la couche supérieure

The Output Layer: cette couche nous donnera une information sur l'échantillon que nous avons donné à notre réseau de neurones. Dans notre cas, il s'agirait de prédire l'adresse IP de destination. Comme notre simulation de flux de données a engagé 16 adresses IP, le nombre de neurones dans cette couche est de 16.

The Hidden Layer: les couches hidden peuvent être composées d'une ou de plusieurs sous-couches dont les nombres et les hyperparamètres ont été déterminés par la recherche de grille. Dans notre cas, il contient deux couches.

À ce stade, nous avons construit notre modèle et prétraité nos données. Il ne reste plus qu'à compiler le modèle, à l'entraîner et à le sauvegarder en utilisant respectivement les fonctions précédemment définies `compile()`, `model.fit()` et `model.save()`. L'évaluation de notre modèle à l'aide de données de validation à un taux de précision de 72,7%.

Approche SDN intelligente proposée

Approche ISDN

Tests des Nouvelles règles API Northbound C-SDN



A. Règle ACL : Liste de contrôle d'accès

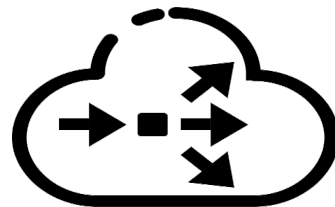


deny

allow



B. Règle d'équilibrage de charge à partir d'une certaine valeur de bande passante



waypoint

Approche SDN intelligente proposée

Approche ISDN

Tests des Nouvelles règles API Northbound C-SDN



C. Règle pour l'allocation d'une grande partie de la bande passante pour des adresse IP bien déterminés.

RateLimit



D. Règle qui fixe un temps de latence maximal

latency

Conclusion Générale

- En se basant sur notre **modèle ML d'apprentissage par renforcement** → un contrôleur SDN sera capable de **prédire les meilleurs Vswitchs**, le plus optimal et performant en termes de Qod.
- En se basant sur la technique de l'apprentissage supervisé → un nouveau nœud : l'**orchestrateur intelligent pourra apprendre et prédire** via notre modèle d'apprentissage supervisé **les meilleurs contrôleurs SDN** en terme de performance à tout instant ou de façon dynamique.
- **En programmant des règles au niveau de l'interface API Sud et Nord** des contrôleurs SDN nous pouvant **ajouter du comportement (Behavior)** à la demande selon notre Data Lake pour n'importe quelle source de données.



Merci pour votre attention

Imène ELLOUMI ZITOUNA