27 novembre 2018

FIGURE 1

0.1 La détection des anomalies

0.2 A savoir

Les entrées de l'algorithme de la détection est un ensemble de traceroutes. Un traceroute (Traceroute) est un ensemble de sauts (Hop). Chaque saut est décrit par un ensemble de signaux (Signal).

Paramètres de l'algorithme de la détection

- Objectif : suivre l'évolution du délais d'un lien au cours du temps en suivant son RTT différentiel par période du temps (*timeWindow*).
- Entrées : l'ensemble des traceroutes stockés dans un fichier, date de début de l'analyse *start*, date du fin de l'analyse *end*, lien à analyser (*link*) et la fenêtre de l'analyse (*timeWindow*).
- Sorties : les dates pendant lesquelles des anomalies ont été détectées.

Soient d_1,d_2, ..., d_N les périodes entre *start* et *end* où d_{i+1} - d_i = d_{j+1} - d_j =*step* pour tout i et j dans$[1, N]$

0.3 Etapes :

1. Trier les traceroutes à analyser par *timeWindow*. En effet, chaque d_i est associé à un ensemble de traceroutes ayant été effectués entre d_i et $d_i + step$ [1].

Les opérations suivantes (2 à 6) concernent les traceroutes par tout d_i.

2. Vérification de la validité de chaque traceroute du chaque d_i. Ces vérifications reprennent les points suivants :

- élimination des traceroutes échoués complètement ;
- élimination du signal contenant une adresse IP privée ;
- élimination du signal qui ne contient pas un RTT ou celui qui contient un RTT négatif ;
- élimination du signal échoué.

1. Chaque traceroute reprend le temps pendant lequel il était effectué.

3. Calcul de la médiane des RTTs par saut , autrement dit, pour tout saut d'un traceroute, on calcul la médiane des RTTs par adresse IP. Soit le saut $h = \{s\}$ où s est un objet Signal, mediane (h)= $\{median(s)\}$ pour tout signal s ayant la même adresse IP. Ainsi, le saut du traceroute est reconstruit en regroupant les signaux par adresse IP et ensuite en calculant leurs RTTs.

4. Inférence des liens topologiques par traceroute. Un lien topologique est formé par chaque deux routeurs consécutifs repris par l'utilitaire traceroute. Ce sont les deux routeurs des deux sauts consécutifs. De manière générale, la figure 2 illustre la constitution des liens par traceroute. RAi avec $i \in [1, N]$ est l'ensemble des routeurs pour le saut A et RBj avec $j \in [1, M]$ est l'ensemble des routeurs pour le saut B, avec N et M deux entiers.

Ainsi, les liens construits sont ceux partant de tout RAi vers tout RBj, où A et B sont deux sauts consécutifs. A l'issue de cette étape, pour tout traceroute, on obtient la liste des liens possibles tout en reprenant des informations générales de la requête traceroute.

5. Caractérisation des liens avec leur RTTs différentiels. A cette étape, on calcul le RTT différentiel d'un lien en calculant la différence entre les RTTs [2] des deux routeurs du lien en question. En plus du RTT différentiel, on note aussi la sonde Atlas ayant effectué la requête traceroute où le lien a été identifié.

6. Fusion des informations d'un lien. Etant donné qu'un lien (IP1, IP2) peut être identifié plusieurs fois pendant un même timeWindow d'une part, et le lien (IP2,IP1) est similaire [3] au lien (IP1, IP2) d'autre part, la fusion permet de construire une nouvelle distribution des RTTs différentiels caractérisant le lien (IP1, IP2) qui reprend les RTTs différentiels du (IP1, IP2) et du (IP2, IP1).

A la fin de l'étape 6, tous les traceroutes sont analysés tout en identifiant leurs liens, et ce par timeWindow. A présent, l'objectif c'est d'identifier les dates pendant lesquelles des anomalies ont été détectées. Pour ce faire, l'idée du travail de référence c'est de conserver, pour un lien donné, une référence du RTT différentiel médian qui sera d'abord comparée avec la médiane courante du RTT différentiel et ensuite mettre à jour cette référence tout au long de la période de l'analyse.

7. Calcul de la médiane et de l'intervalle de confiance courant du lien analysé.

8. Mise à jour de la médiane et de l'intervalle de référence du lien analysé

9. La détection des anomalies en comparant l'état du lien courant avec l'état de la référence pour ce même lien.

2. C'est la médiane calculée à l'étape 3.
3. La similarité est mesurée par le RTT différentiel.

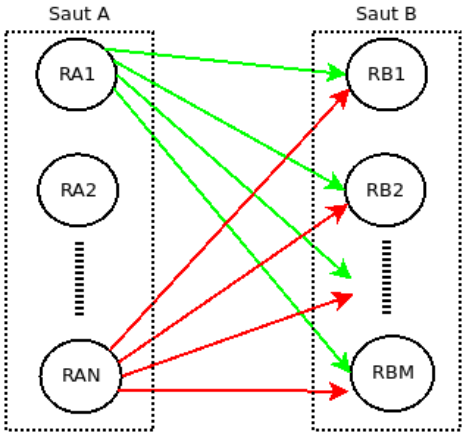

FIGURE 2

0.4 Vue globale des étapes de la détection des anomalies

La figure 3 présente la succession des étapes de la détection des anomalies dans les délais d'un lien donné. Comme complément à ces étapes, on présente les différentes classes permettant de manipuler les données tout au long du processus de l'analyse.

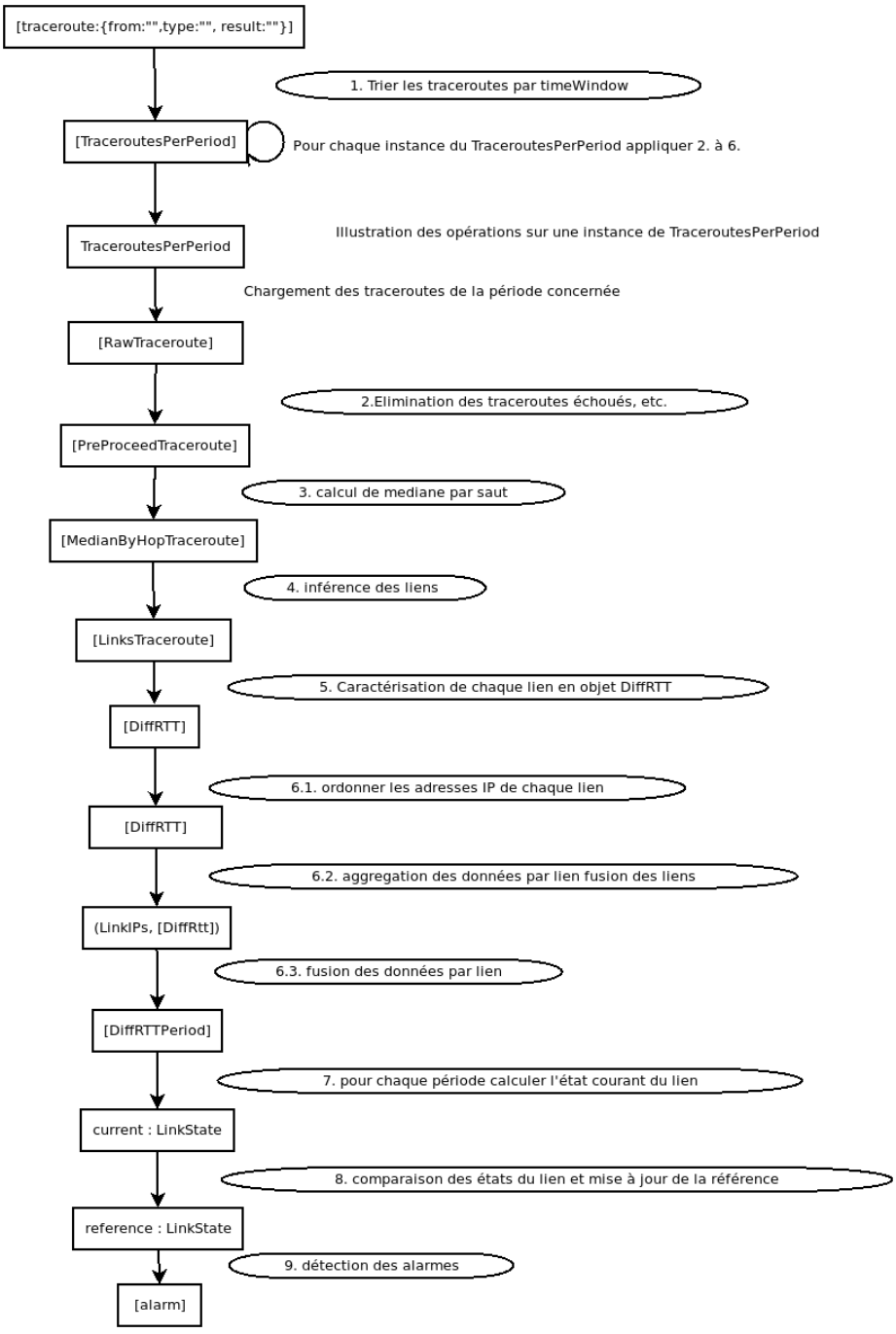
[traceroute:{from:"",type:"", result:""}]
1. Trier les traceroutes par timeWindow
[TraceroutesPerPeriod]
Pour chaque instance du TraceroutesPerPeriod appliquer 2. à 6.
TraceroutesPerPeriod
Illustration des opérations sur une instance de TraceroutesPerPeriod
Chargement des traceroutes de la période concernée
[RawTraceroute]
2.Elimination des traceroutes échoués, etc.
[PreProceedTraceroute]
3. calcul de mediane par saut
[MedianByHopTraceroute]
4. inférence des liens
[LinksTraceroute]
5. Caractérisation de chaque lien en objet DiffRTT
[DiffRTT]
6.1. ordonner les adresses IP de chaque lien
[DiffRTT]
6.2. aggregation des données par lien fusion des liens
(LinkIPs, [DiffRtt])
6.3. fusion des données par lien
[DiffRTTPeriod]
7. pour chaque période calculer l'état courant du lien
current : LinkState
8. comparaison des états du lien et mise à jour de la référence
reference : LinkState
9. détection des alarmes
[alarm]

FIGURE 3

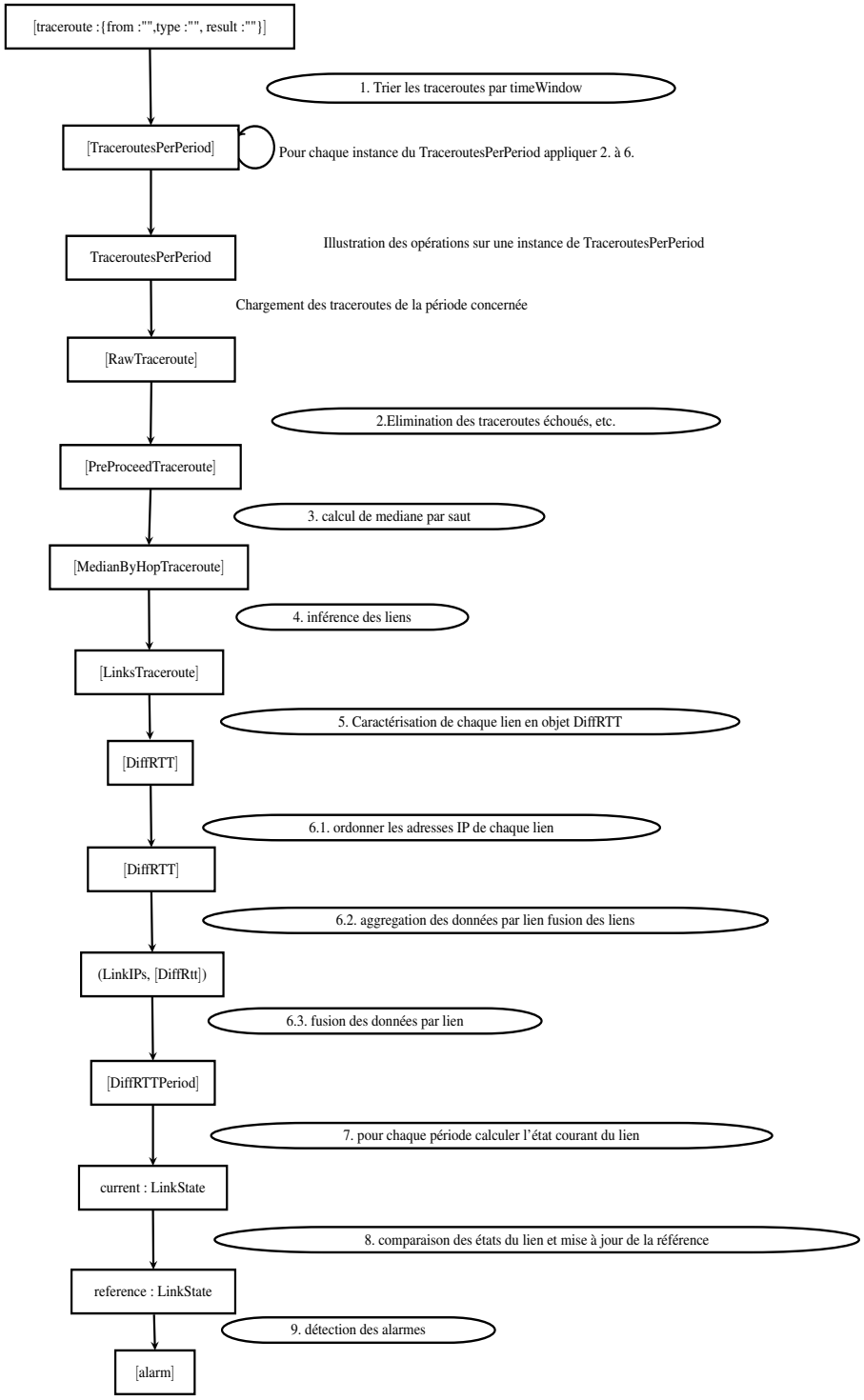
[traceroute :{from :"",type :"", result :""}]
1. Trier les traceroutes par timeWindow
[TraceroutesPerPeriod]
Pour chaque instance du TraceroutesPerPeriod appliquer 2. à 6.
TraceroutesPerPeriod
Illustration des opérations sur une instance de TraceroutesPerPeriod
Chargement des traceroutes de la période concernée
[RawTraceroute]
2.Elimination des traceroutes échoués, etc.
[PreProceedTraceroute]
3. calcul de mediane par saut
[MedianByHopTraceroute]
4. inférence des liens
[LinksTraceroute]
5. Caractérisation de chaque lien en objet DiffRTT
[DiffRTT]
6.1. ordonner les adresses IP de chaque lien
[DiffRTT]
6.2. aggregation des données par lien fusion des liens
(LinkIPs, [DiffRtt])
6.3. fusion des données par lien
[DiffRTTPeriod]
7. pour chaque période calculer l'état courant du lien
current : LinkState
8. comparaison des états du lien et mise à jour de la référence
reference : LinkState
9. détection des alarmes
[alarm]

FIGURE 4