TELECOM Paris

IP PARIS

RES 101

L’architecture de l’Internet

- Routage et Relayage
- Les algorithmes de routage classiques

TELECOM Paris

IP PARIS

Routing & *Forwarding*

Le **routage** et le *relayage*

(How to know where to send a packet?)

Routing & Forwarding

- Qu'est-ce que c'est que le routage ?
  - Processus pour échanger les information sur la topologie entre les nœuds d'un réseau .
  - Il fournit les moyens pour calculer le "meilleur" chemin vers un nœud destination.
  - Il utilise une structure de données : la table de routage (routing table).
    - RIB - Routing Information Base

- Qu'est-ce que c'est que le relayage ?
  - Processus qui redirige un paquet entrant un port / interface (**input**) vers un autre port / interface (**output**)
  - Il utilise les informations obtenues grâce aux fonctions de routage.
  - Il utilise une structure de données : la table de forwarding (forwarding table).
    - FIB - Forwarding Information Base

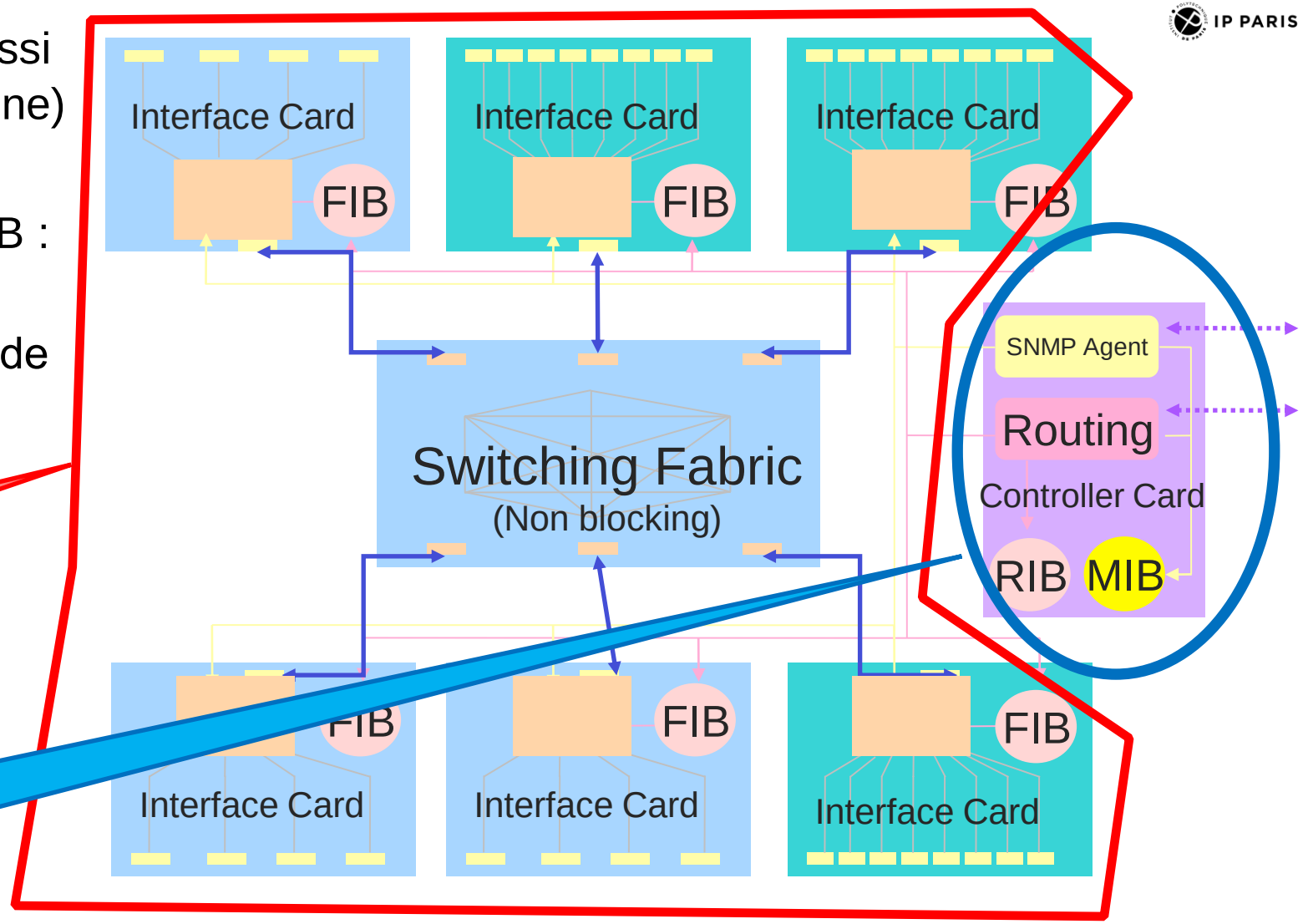
Routing & Forwarding : en gros

- Le routage (**routing**) permet de remplir la **table de routage** (routing table) qui contient les information sur quel destination est en quelle direction
  - Le routage est traité dans le prochain chapitre

- Le relayage (**forwarding**) utilise le contenu de la table de routage pour envoyer les paquets dans la bonne direction
  - Le relayage est traité dans ce chapitre

L'architecture d'un routeur | composants hardware

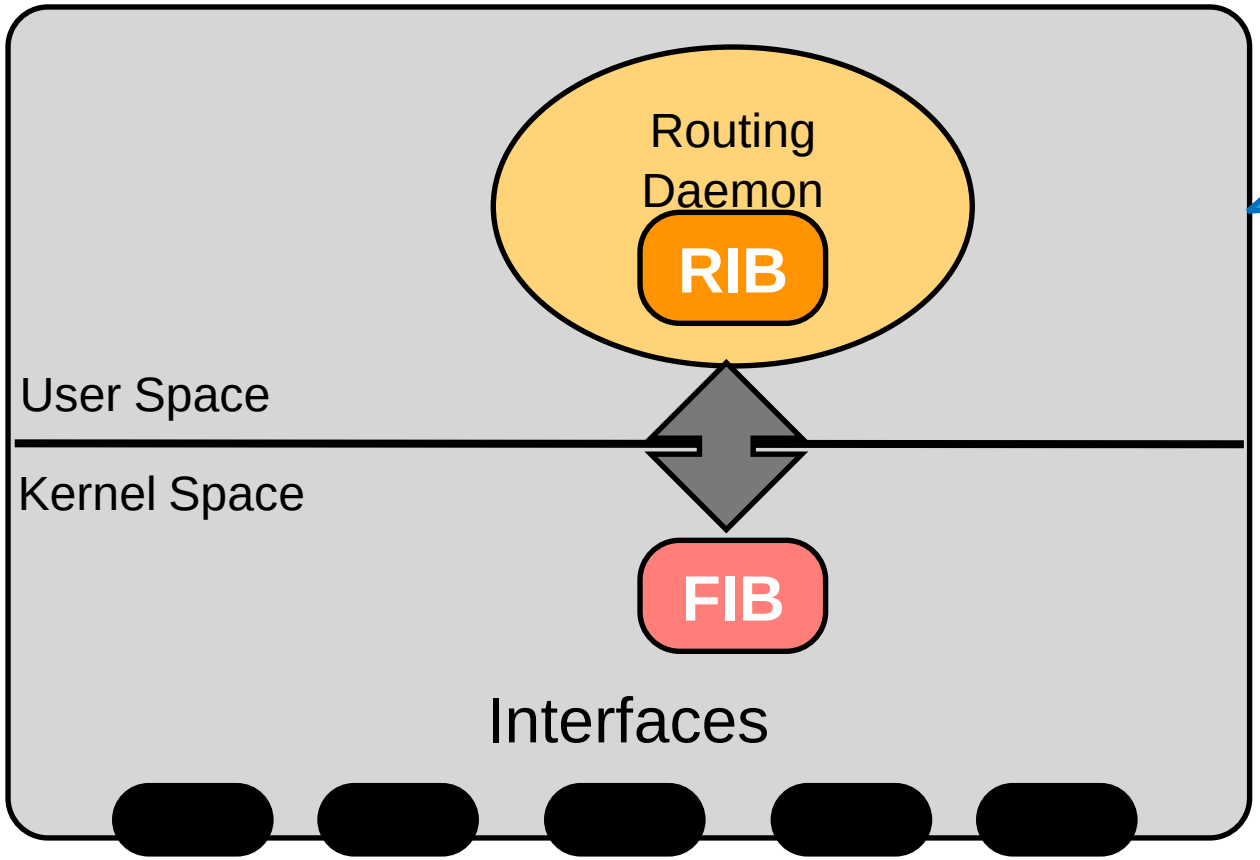
- Les interfaces, ou NICs sont aussi appelés *line cards* (cartes de ligne)
- 1 Line card → plusieurs ports
- Chaque line card contient un FIB : **Forwarding Information Base**
- Le FIB n'a pas les informations de routage complètes, mais seulement ce qui sert pour le forwarding



Data Plane
(All operations in hardware)

Control Plane
(Operations de management, quelque traitement de paquets specifiques (in software).
Jusqu'à 1000x plus lent que le data plane)

Software Router Architecture (Linux/FreeBSD/ etc..)



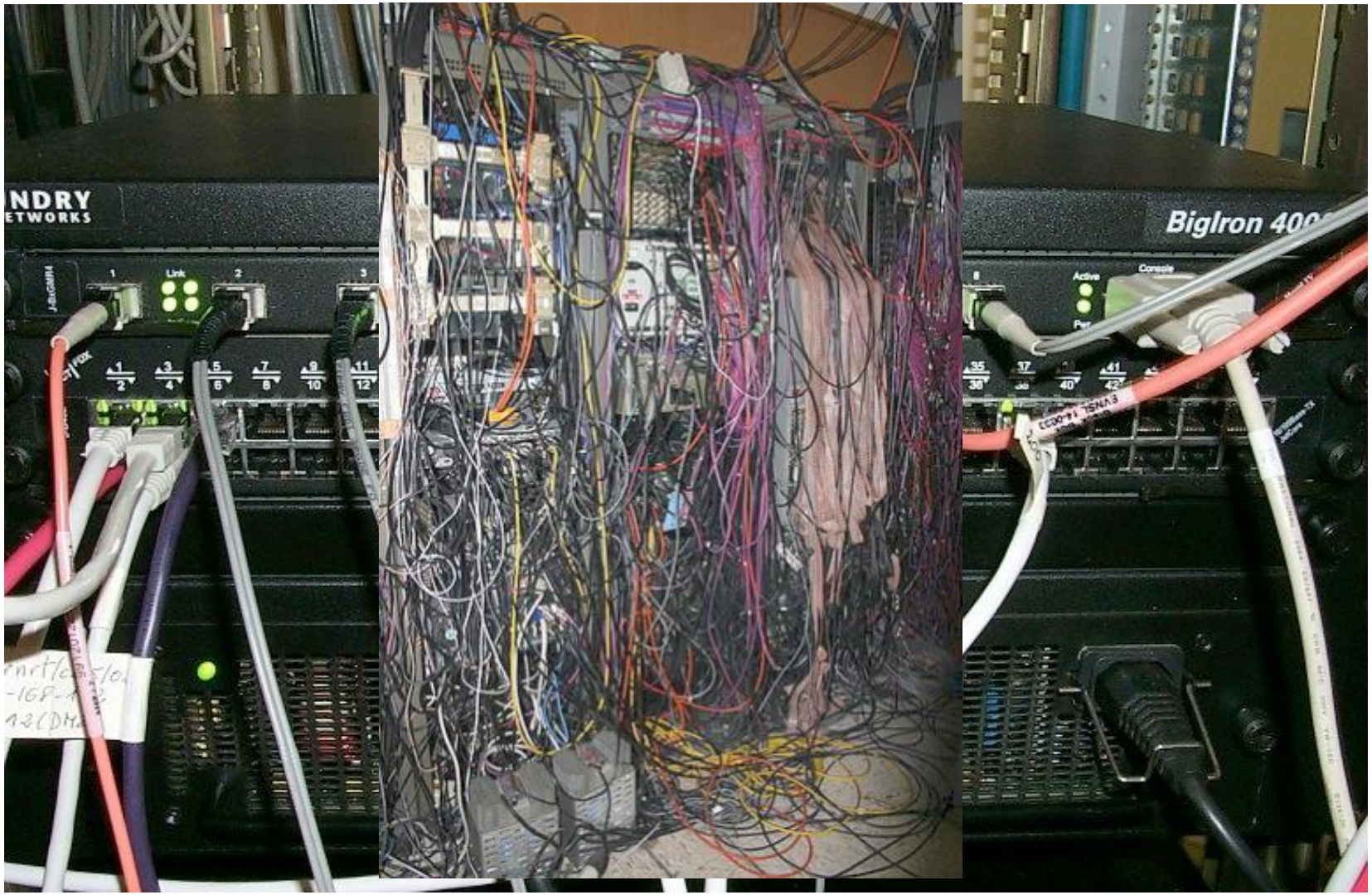
Control Plane

(Le management et quelques cas particulière de traitement de paquets sont effectués en software.
Il peut être même plus rapide que le data plane)

Data Plane

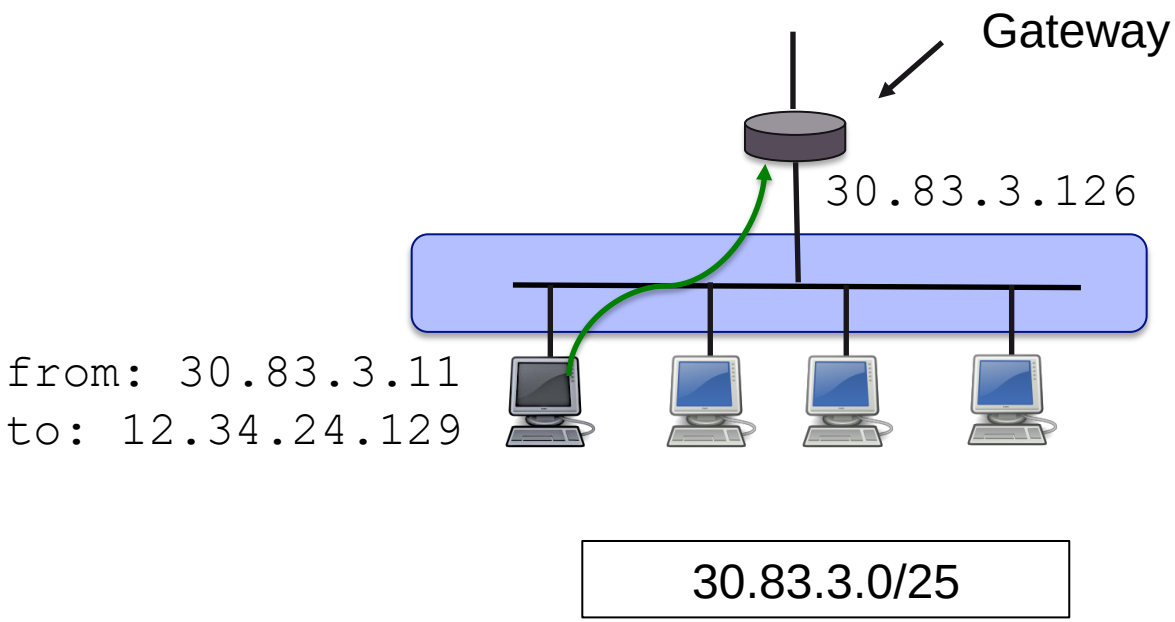
Relayage logiciel, mais à l'intérieur du OS.

Routeurs....

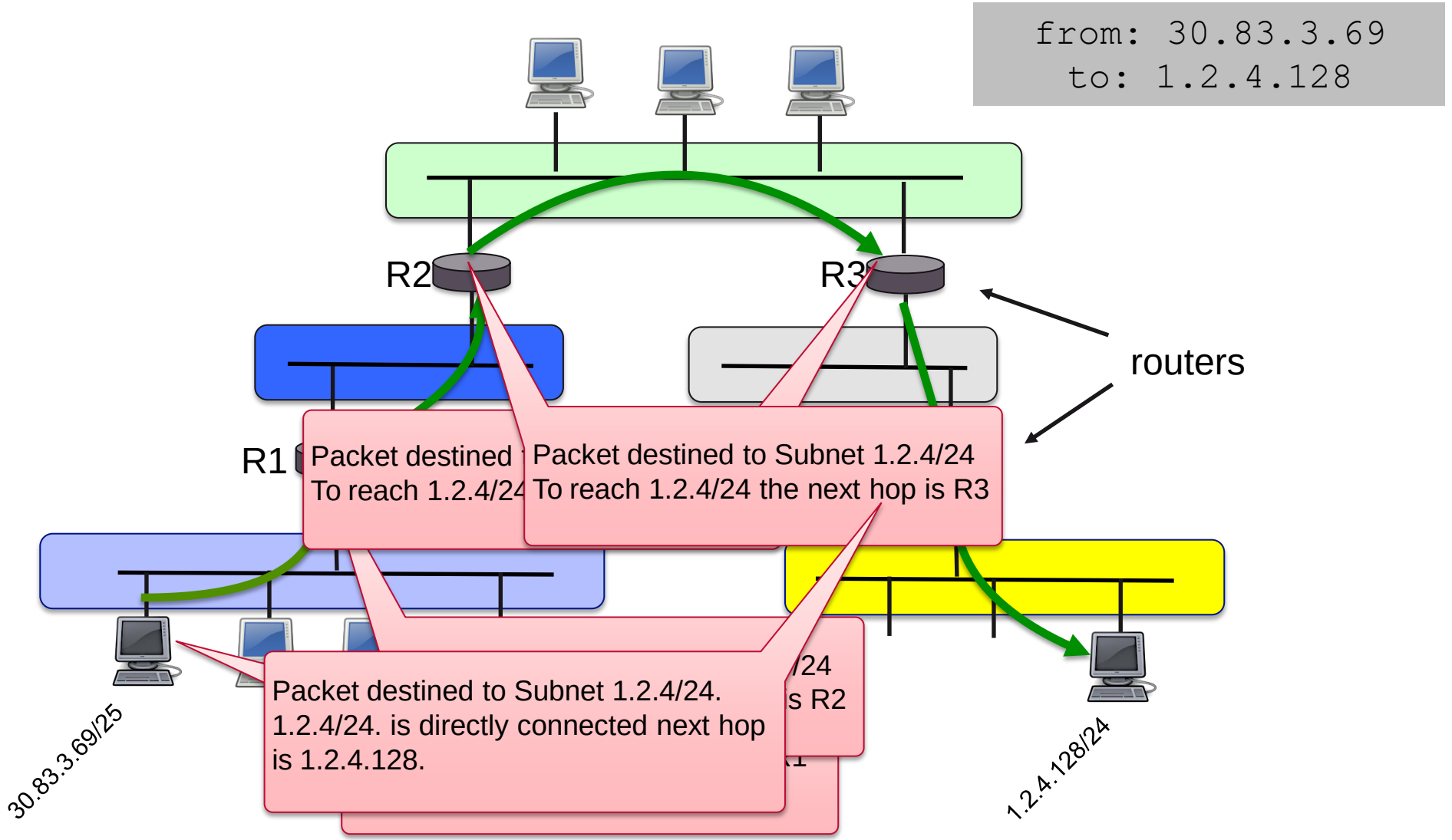


Gateway

- Définition
  - Un GW est un routeur qui connecte deux ou plusieurs (sous)réseaux

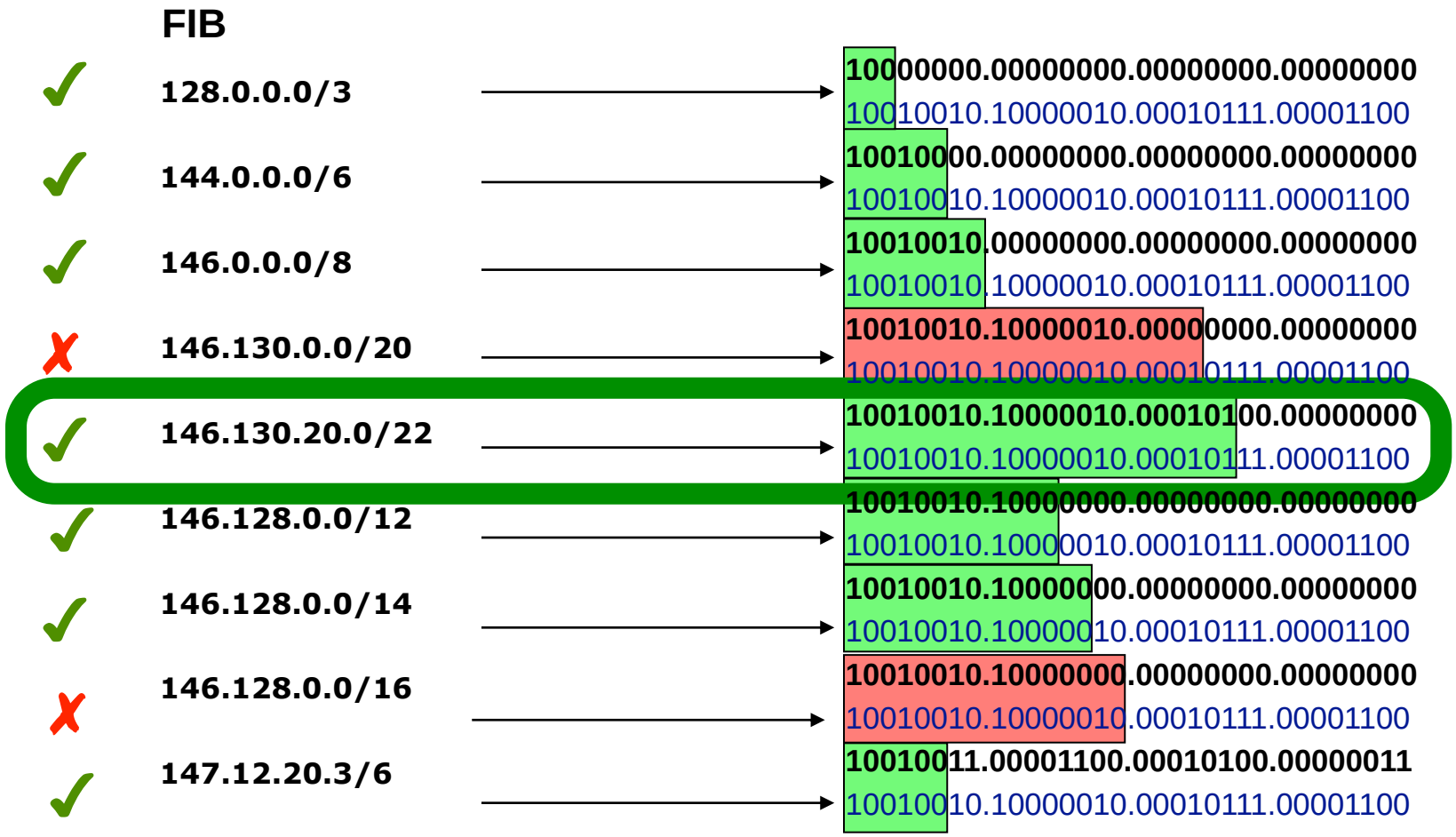


Following a path…



Longest Prefix Match

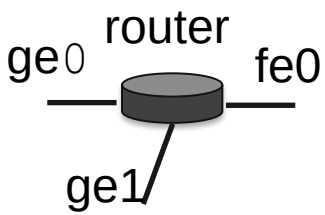
@Dest: 146.130.23.12 **10010010.10000010.00010111.00001100**



LPM & Routing Table

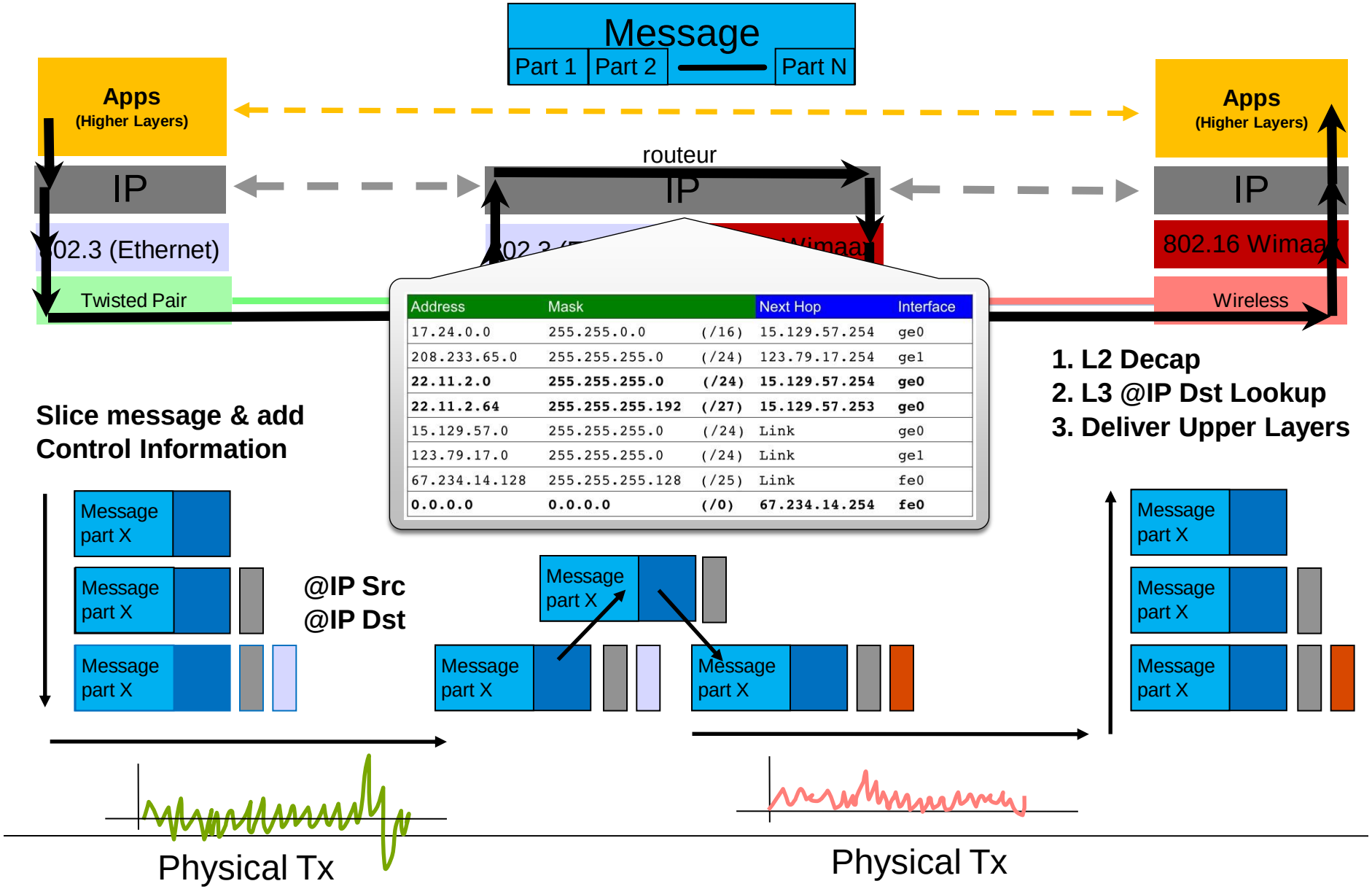
Longest-Prefix Match Rule

E.g. 22.11.2.65

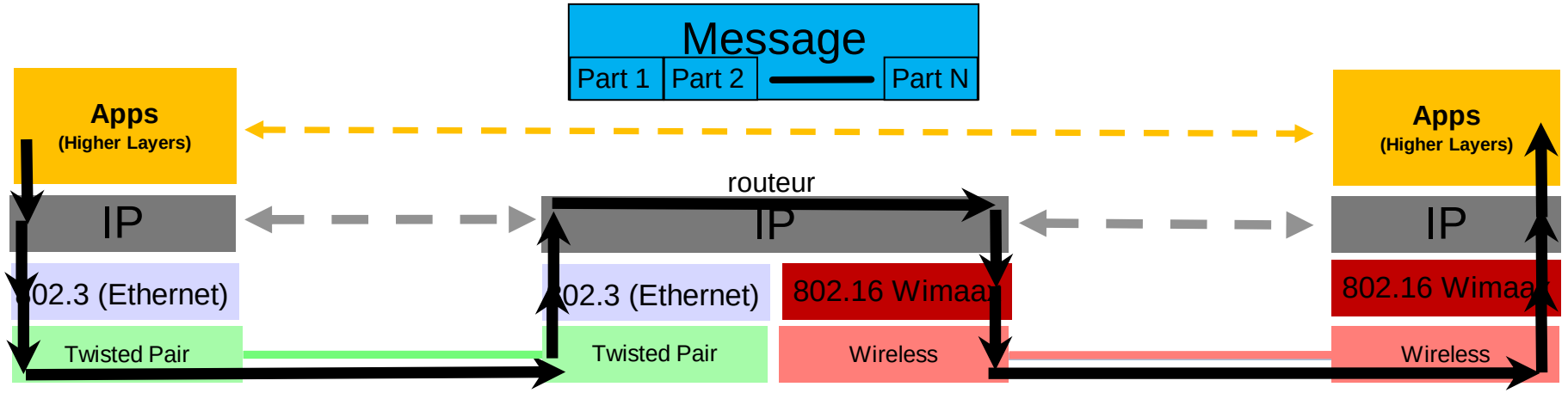


Address	Mask		Next Hop	Interface
17.24.0.0	255.255.0.0	(/16)	15.129.57.254	ge0
208.233.65.0	255.255.255.0	(/24)	123.79.17.254	ge1
22.11.2.0	**255.255.255.0**	**(/24)**	**15.129.57.254**	**ge0**
22.11.2.64	**255.255.255.192**	**(/27)**	**15.129.57.253**	**ge0**
15.129.57.0	255.255.255.0	(/24)	Link	ge0
123.79.17.0	255.255.255.0	(/24)	Link	ge1
67.234.14.128	255.255.255.128	(/25)	Link	fe0
0.0.0.0	**0.0.0.0**	**(/0)**	**67.234.14.254**	**fe0**

Transmission bout-en-bout - Les opérations dans un routeur



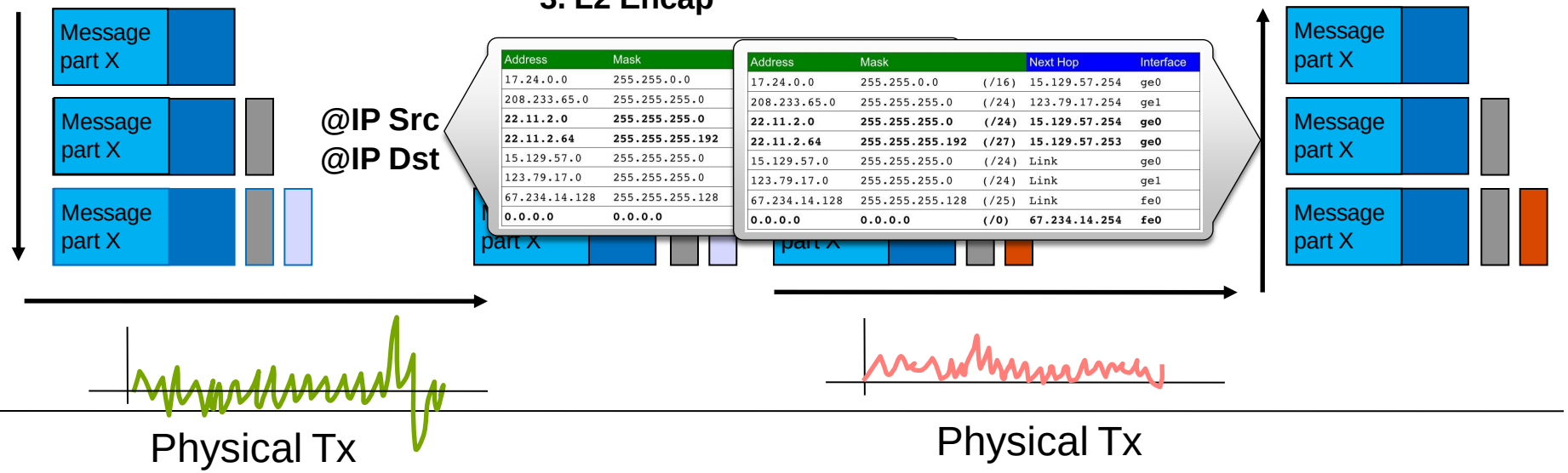
Transmission bout-en-bout – Source/Dest.



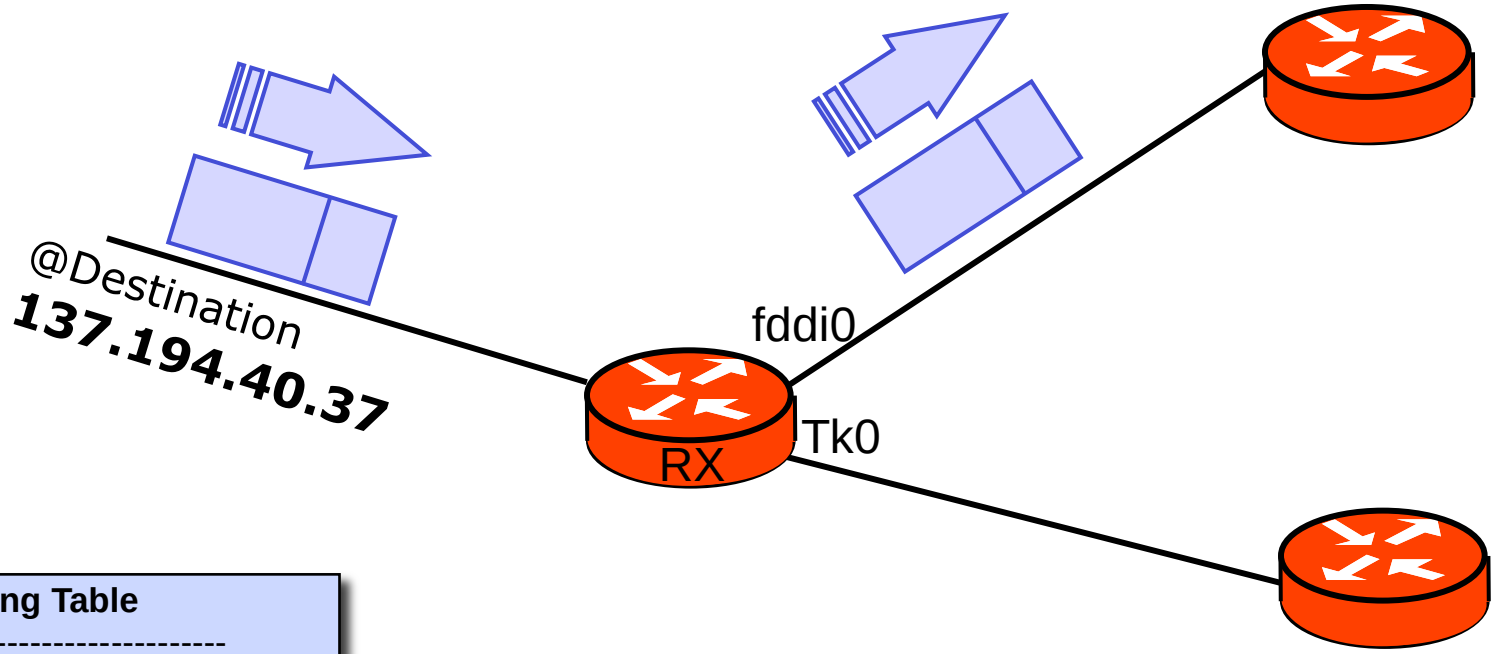
Slice message & add Control Information

1. L2 Decap
2. L3 @IP Dst Lookup
3. L2 Encap

1. L2 Decap
2. L3 @IP Dst Lookup
3. Deliver Upper Layers



Relayage vers un **next hop**



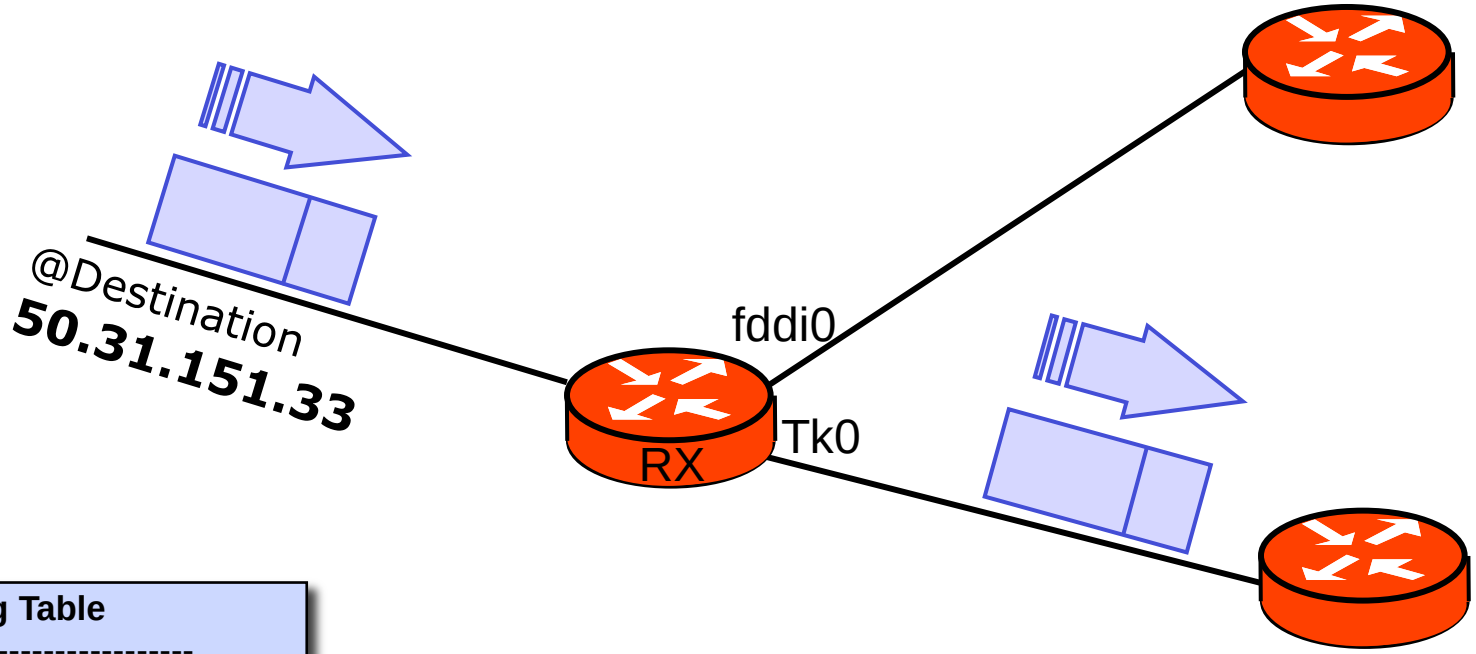
Routing Table

137.194.10.0/24	R5	Tk0
137.194.20.0/24	local	Tk0
137.194.30.0/24	local	Eth0
137.194.40.0/24	R4	fddi0
137.194.50.0/24	local	fddi0
137.194.0.0/16	G2	fddi0
200.0.0.0/10	G2	fddi0
Default Route (0/0)	G1	fddi0

137.194.40.0/24	R4	fddi0
137.194.0.0/16	G2	fddi0
Default Route (0/0)	G1	fddi0

Next Hop: R4
Output Port: fddi0

Relayage vers la **Default Route** | la route par défaut



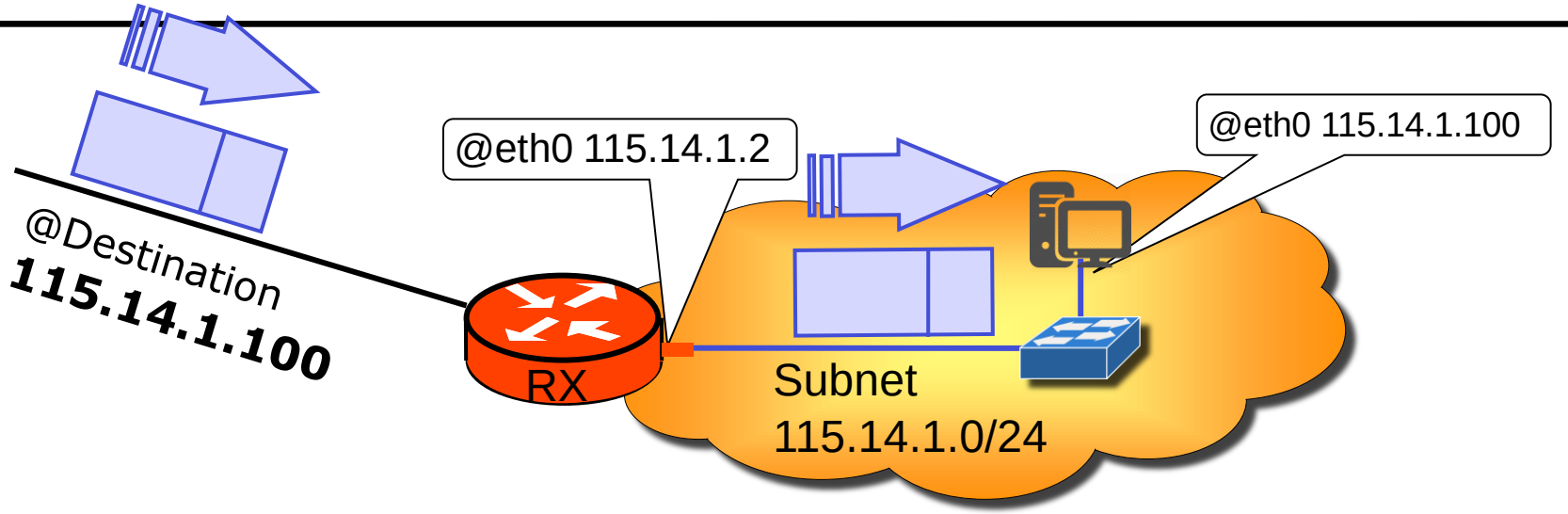
Routing Table

137.194.10.0/24	R5	Tk0
137.194.20.0/24	local	Tk0
137.194.30.0/24	local	Eth0
137.194.40.0/24	R4	fddi0
137.194.50.0/24	local	fddi0
137.194.0.0/16	G2	fddi0
200.0.0.0/10	G2	fddi0
Default Route (0/0)	R2	Tk0

Default Route (0/0) R2 Tk0

Next Hop: R2

Relayage vers une **destination locale**

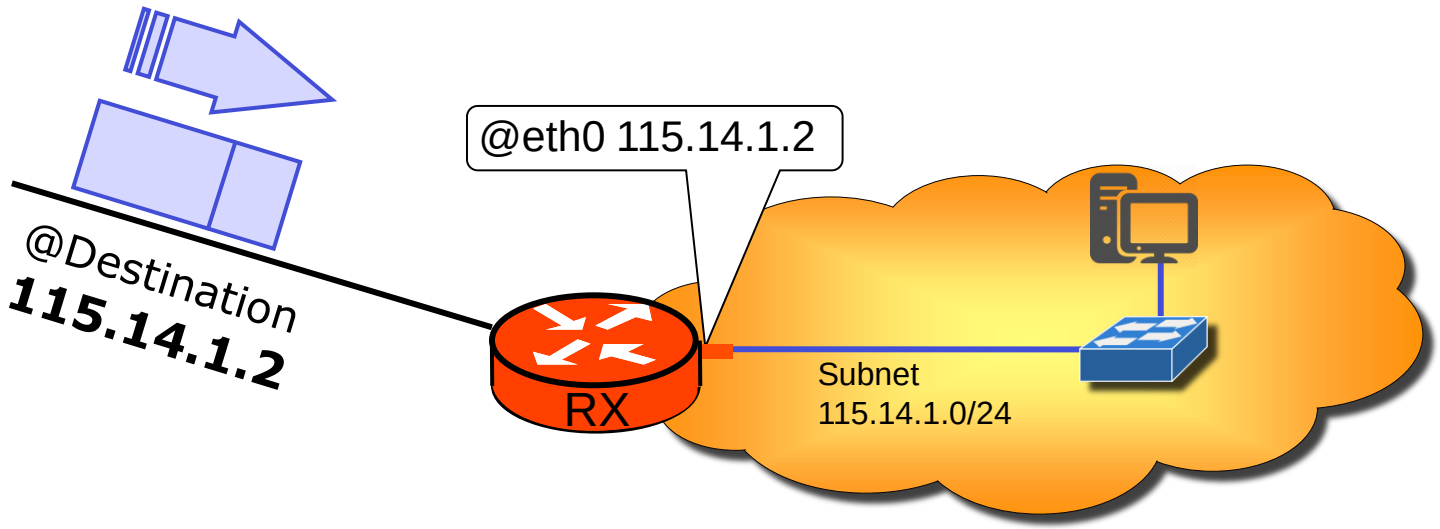


Routing Table		
…		
…		
115.14.1.2/32	local	eth0
115.14.1.0/24		eth0
…		
…		
Default route	G1	fddi0

115.14.1.0/24 eth0

Next Hop: -
La destination est dans le même réseau (l'interface eth0 du routeur).
Le paquet est envoyé directement vers la destination avec encapsulation L2.

Destination locale | Les couches supérieures



Routing Table

…		
…		
115.14.1.2/32	local	eth0
115.14.1.0/24		eth0
…		
…		
Default route	G1	fddi0

115.14.1.2/32	local	eth0
115.14.1.0/24		eth0

Next Hop: - Local!
(Decapsulate and hand packet to transport layer)

Rappelle: Routing & Forwarding

- Qu'est-ce que c'est que le routage ?
  - Processus pour échanger les information sur la topologie entre les nœuds d'un réseau .
  - Il fournit les moyens pour calculer le "meilleur" chemin vers un nœud destination.
  - Il utilise une structure de données : la table de routage (routing table).
    - RIB - Routing Information Base

- Qu'est-ce que c'est que le relayage ?
  - Processus qui redirige un paquet entrant un port / interface (**input**) vers un autre port / interface (**output**)
  - Il utilise les informations obtenues grâce aux fonctions de routage.
  - Il utilise une structure de données : la table de forwarding (forwarding table).
    - FIB - Forwarding Information Base

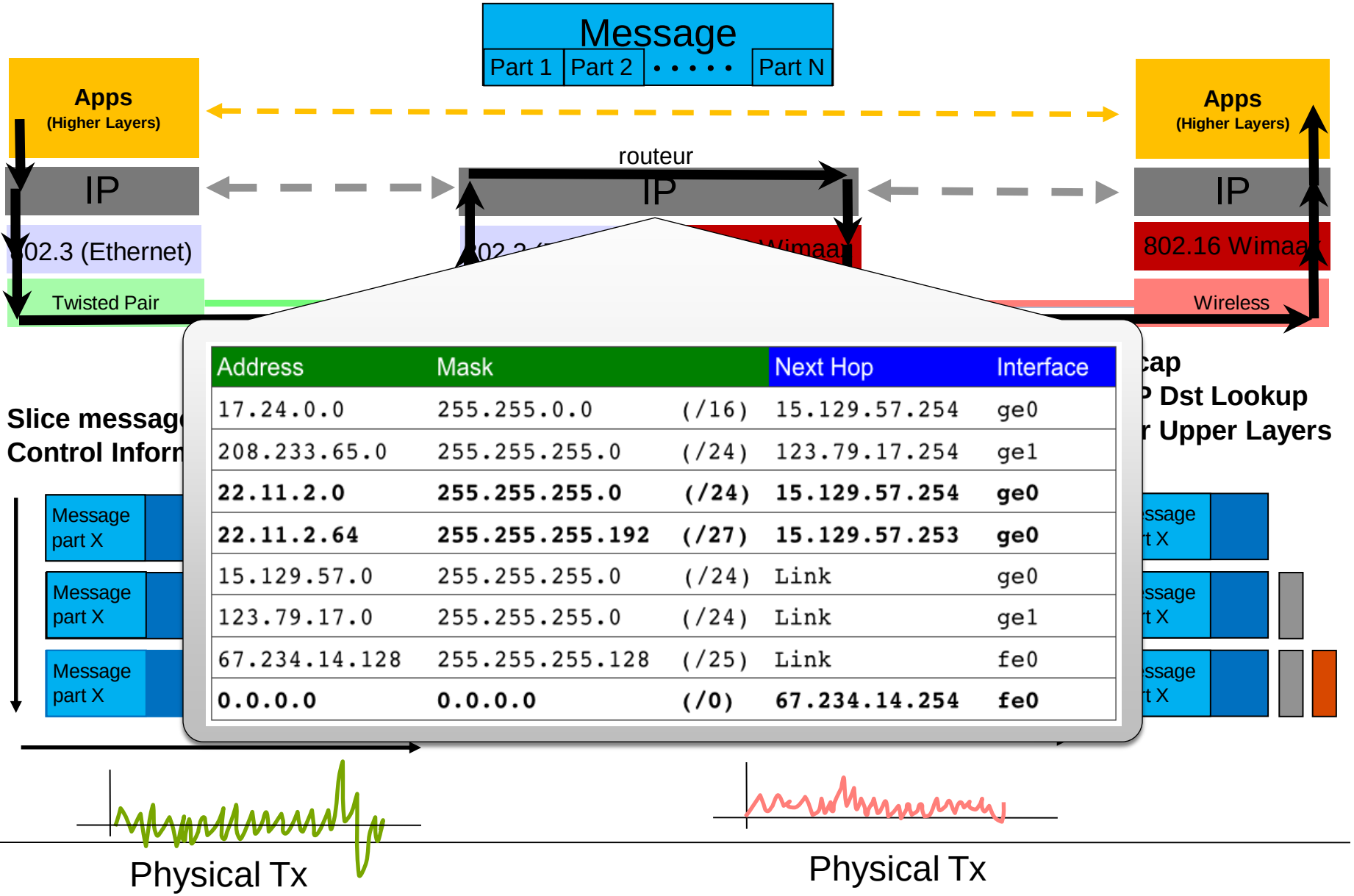
Agenda

- Routage dynamique
- Inter-Domain Routing
  - BGP – Border Gateway Protocol
- Intra-Domain Routing
  - OSPF – Open Shortest Path First
  - RIP – Routing Information Protocol

Routage dynamique

(How to know where to send a packet?)

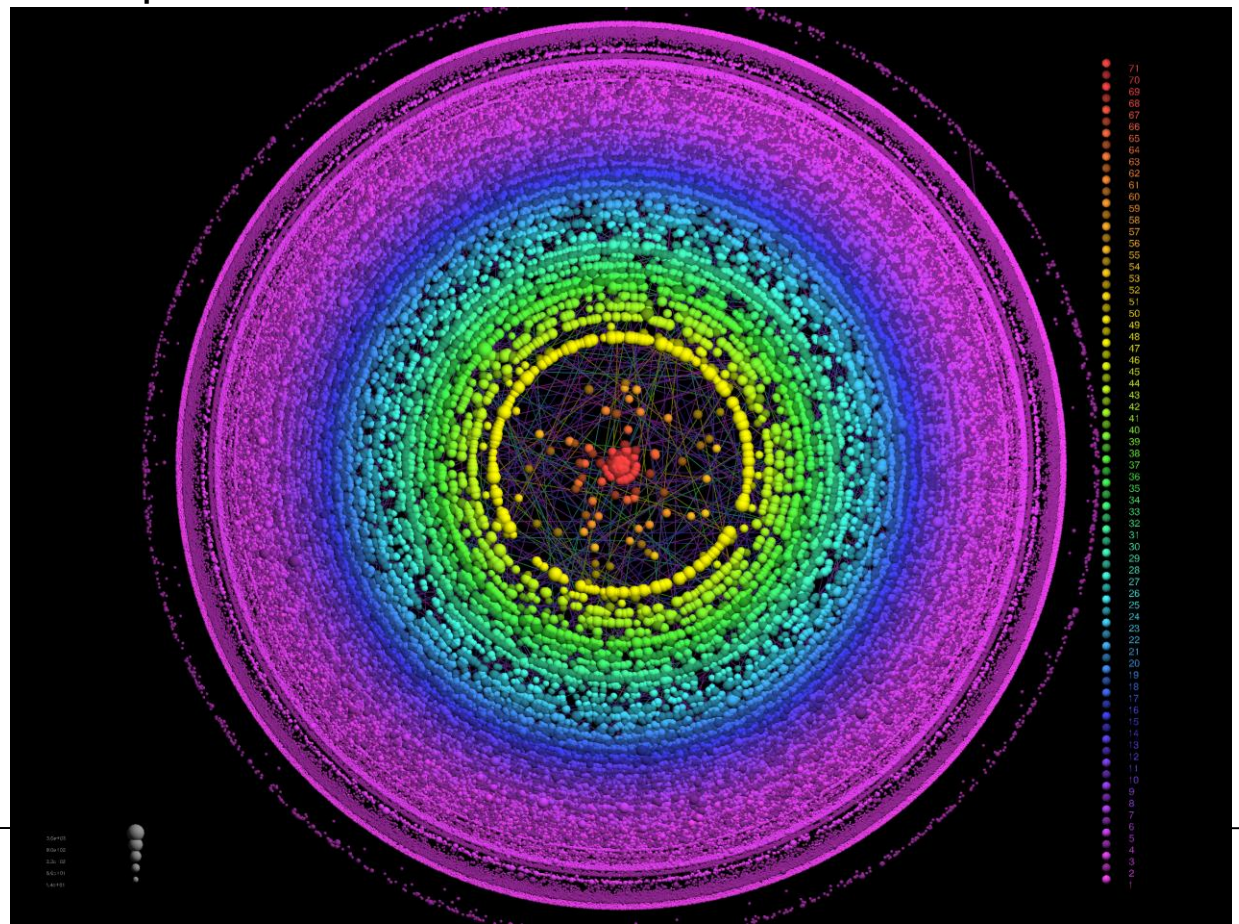
Comment remplir la table de routage ?



Address	Mask		Next Hop	Interface
17.24.0.0	255.255.0.0	(/16)	15.129.57.254	ge0
208.233.65.0	255.255.255.0	(/24)	123.79.17.254	ge1
22.11.2.0	**255.255.255.0**	**(/24)**	**15.129.57.254**	**ge0**
22.11.2.64	**255.255.255.192**	**(/27)**	**15.129.57.253**	**ge0**
15.129.57.0	255.255.255.0	(/24)	Link	ge0
123.79.17.0	255.255.255.0	(/24)	Link	ge1
67.234.14.128	255.255.255.128	(/25)	Link	fe0
0.0.0.0	**0.0.0.0**	**(/0)**	**67.234.14.254**	**fe0**

Routage statique

- Applicable aux petit réseaux
- **Administration intensive –** chaque changement est fait manuellement dans chaque routeur
- Principalement utilisé pour définir la route par défaut
  - 0.0.0.0/0 Next Hop Router



IP Networks Graph
http://www.netdimes.org/community.html

Routage dynamique – les avantages

- Possibilité de détecter automatiquement les changement de topologie et s'adapter
- Scalability
- Récupération algorithmique du meilleur chemin
- Robustness
- Simplicité
- Convergence rapide
- Possibilité de « bidouiller » avec les routes (Traffic Engineering)
    - E.g. quel lien est-il souhaitable pour des raisons d'optimisation ?

Les caractéristiques du routage dynamique

- Ce qu'il fait
  - Mécanismes pour partager la connaissance sur les préfixes IP
    - Et donc, remplir les tables de routage

- Ce qu'il **ne fait pas** !
  - Configurer les adresses IP pour les interfaces réseau
    - On utilise DHCP pour les end-host
    - Adressage statique dans les routeurs
      - **Static ≠ Manually configured**
  - Fournir des routes par défaut
    - Sauf si explicitement configuré
  - Décider les policies pour les liens
    - E.g., quelles informations propager et vers qui
      - Ceci sera fait par d'autres techniques (BGP)
    - Filtrage

L’architecture de l’Internet n’est pas « monolithique »

- Il existes plusieurs ...
  - Routing Protocol
  - Routing Configuration
  - Routing State
  - Routing Management
- ..... tout est distribué dans l’Internet !

- Le système architectural de l’Internet est basé sur plusieurs composant qui sont censés operér de manière consistante (**hopefully**…)
  - On appelle **convergence** quand tous les routeurs partagent les mêmes informations de routage

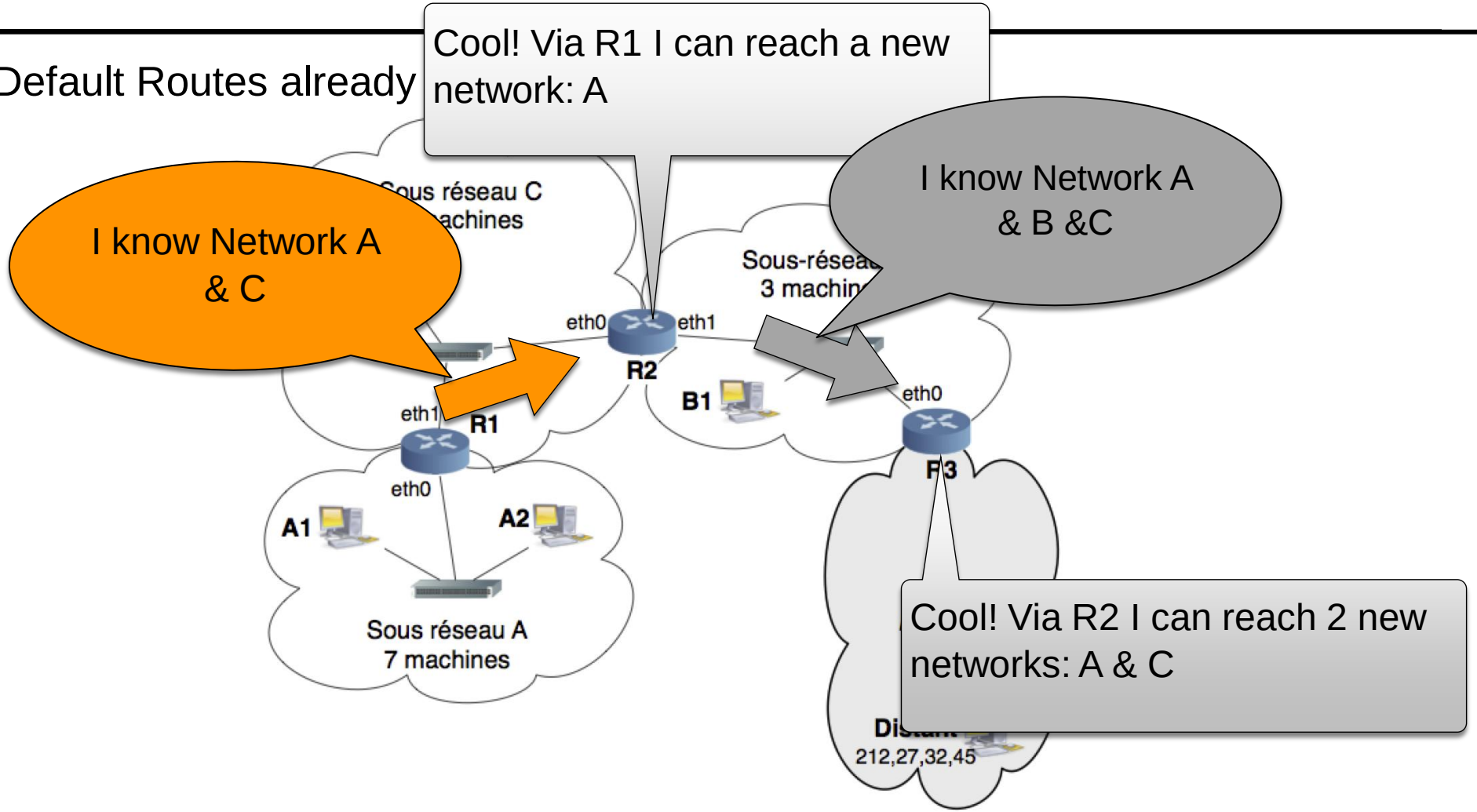
Le principe du routage dynamique : Gossip!

- Tous les systèmes de routage dynamique partagent la même approche :
  - « *I tell you what I know and you tell me what you know!* »

- Tous les systèmes de routage ont les objectives :
  - De créer une vision consistante de l'Internet
  - D'éviter les boucles
  - D'éviter les « trous noirs » (morceaux du réseau non joignable)
  - Trouver des chemins "optimaux" (or "best path")
    - La définition d'optimalité peut varier

Pourquoi partager les informations de routage ?

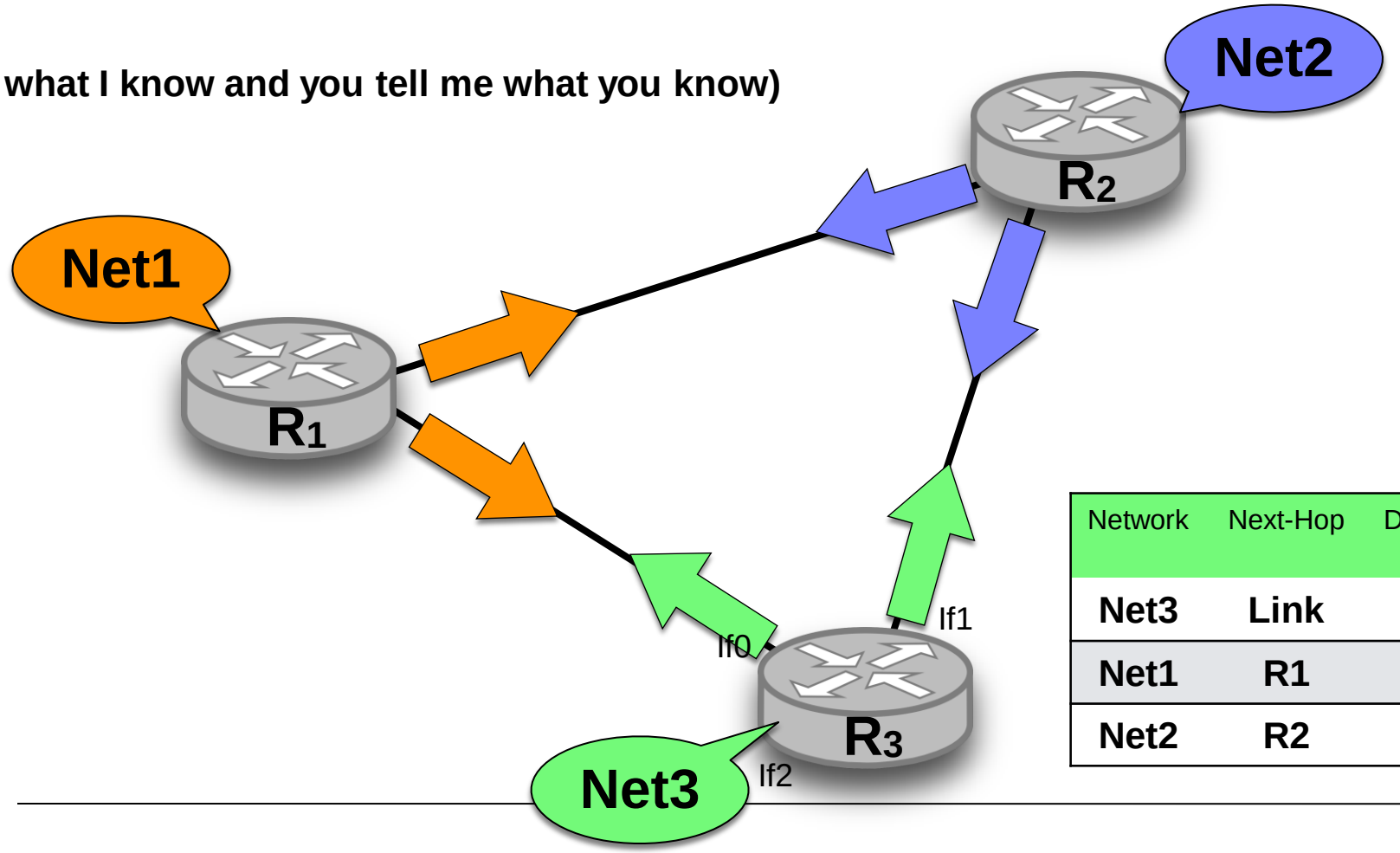
- Assume Default Routes already



Routage dynamique – échange des informations

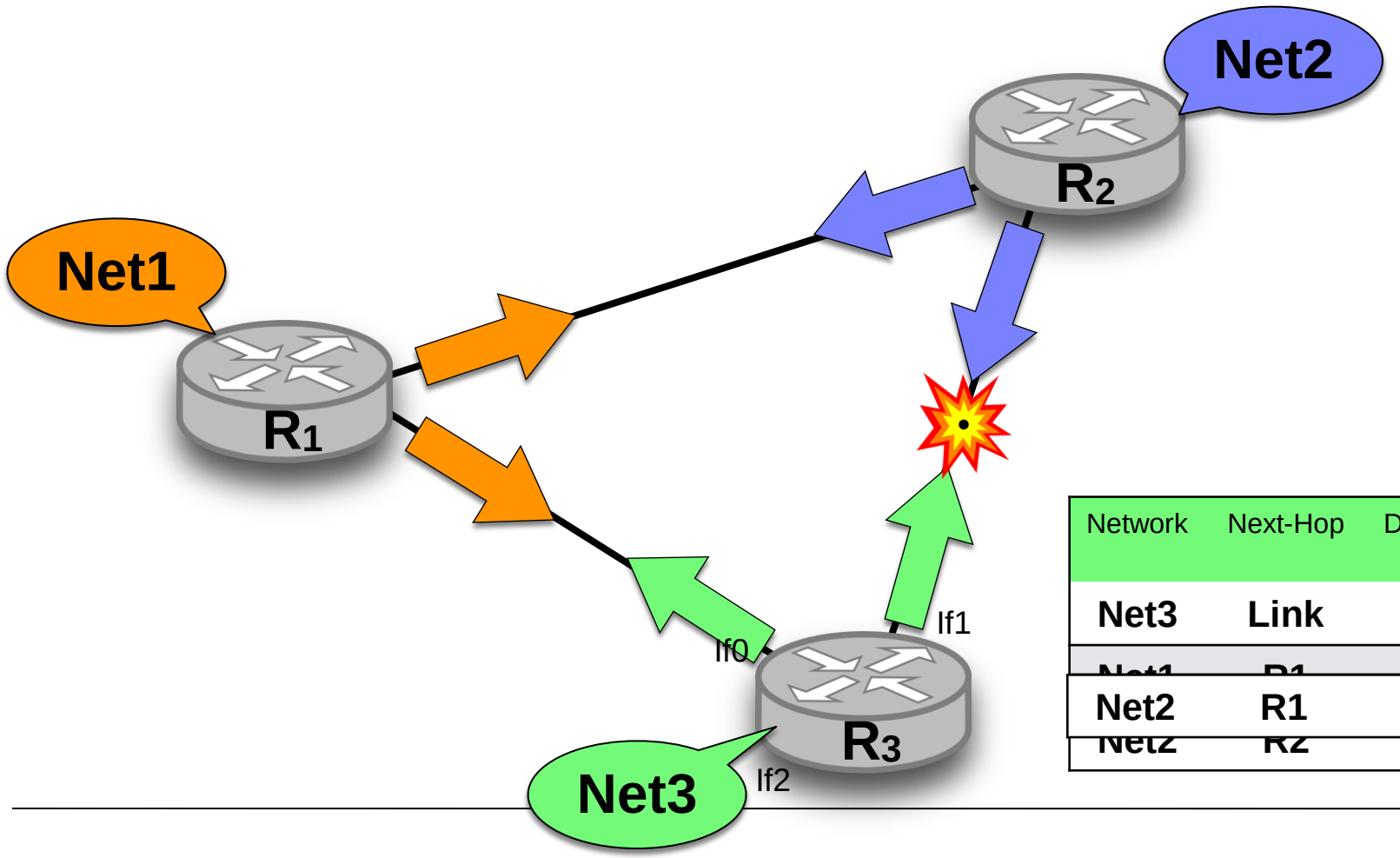
- Detection automatique des connexions (**reachability**)
- Computation automatique du « meilleur » chemin

- **(I tell you what I know and you tell me what you know)**



Network	Next-Hop	Distance	Interface
Net3	**Link**	**0**	**If2**
Net1	**R1**	**1**	**If0**
Net2	**R2**	**1**	**If1**

Routage dynamique : réaction aux défauts



Network	Next-Hop	Distance	Interface
Net3	Link	0	If2
Net2	R1	2	If0

Address Aggregation

10000110.00100011.00000100.00000000

10000110.00100011.00000101.00000000

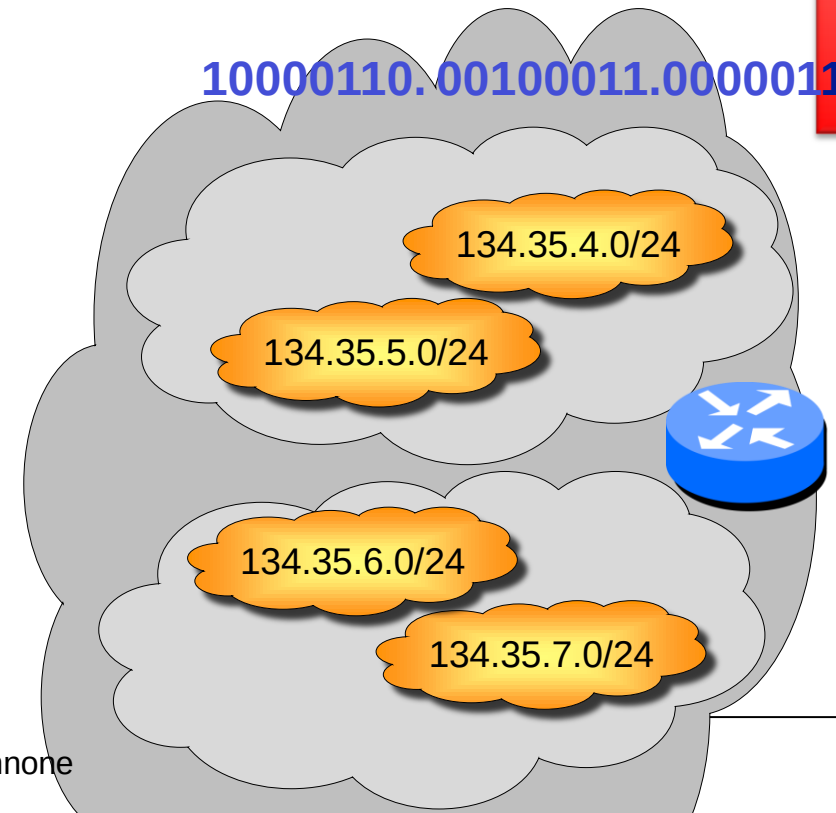
10000110.00100011.00000110.00000000

10000110.00100011.00000111.00000000

10000110.00100011.0000010 0.00000000

10000110.00100011.00000100.00000000

10000110.00100011.00000110.00000000



134.35.4.0/24
134.35.5.0/24
134.35.6.0/23
134.35.7.0/24

134.35.4.0/22

Aggregation…. Yes! But be careful …..

10000110.00100011.00010001.00000000

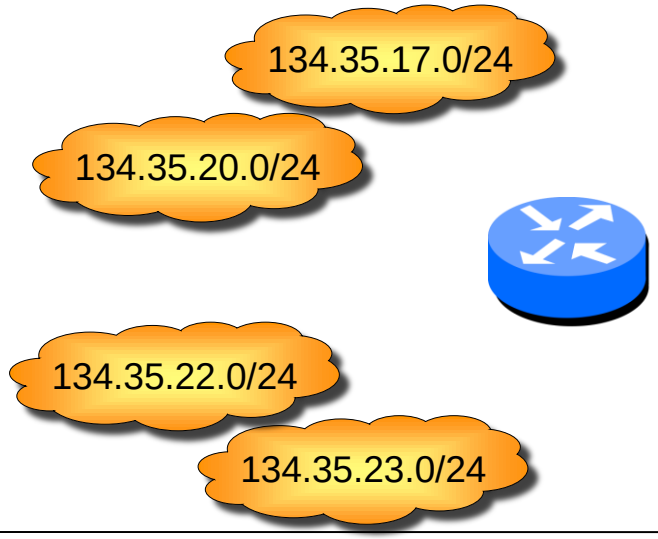
10000110. 00100011.00010100.00000000

10000110. 00100011.00010110.00000000

10000110. 00100011.00010111.00000000

~~10000110.00100011.00010100.00000000~~

~~134.35.20.0/22 ?~~



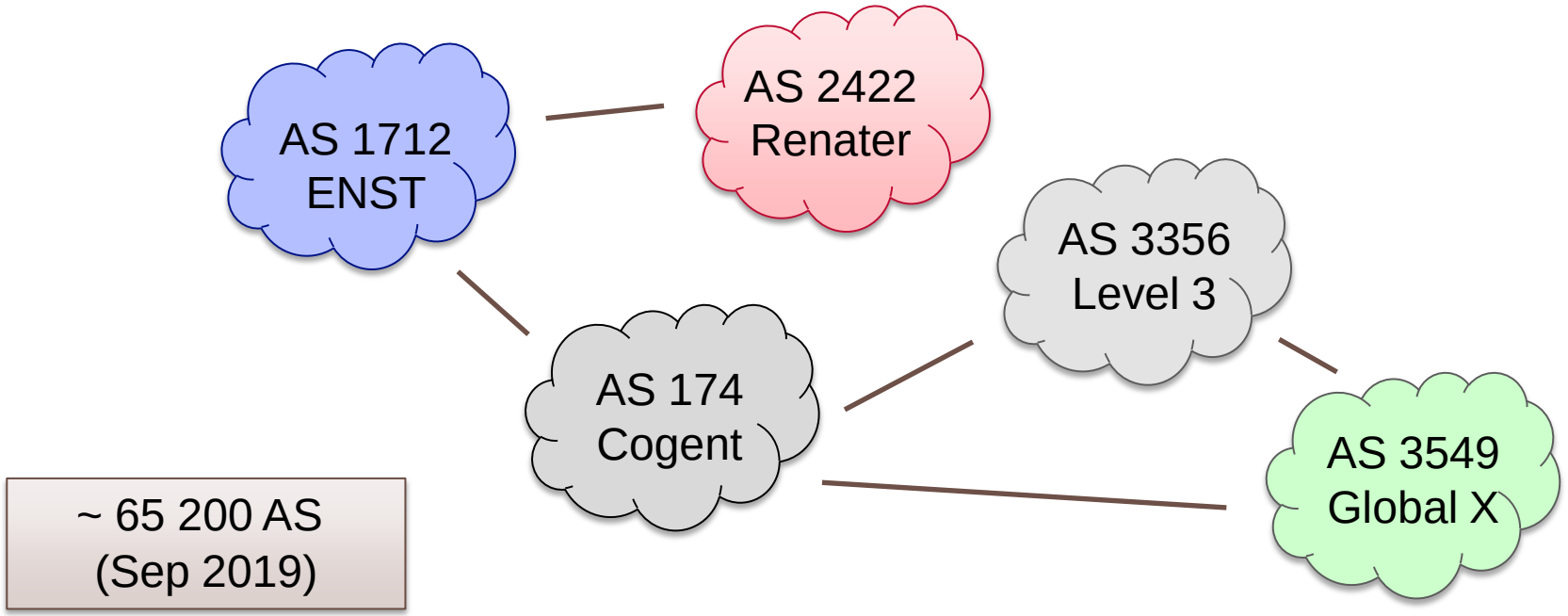
134.35.17.0/24
134.35.20.0/24
134.35.22.0/24
134.35.23.0/24

Le routage entre différents réseaux

(How are address information spread in the Internet?)

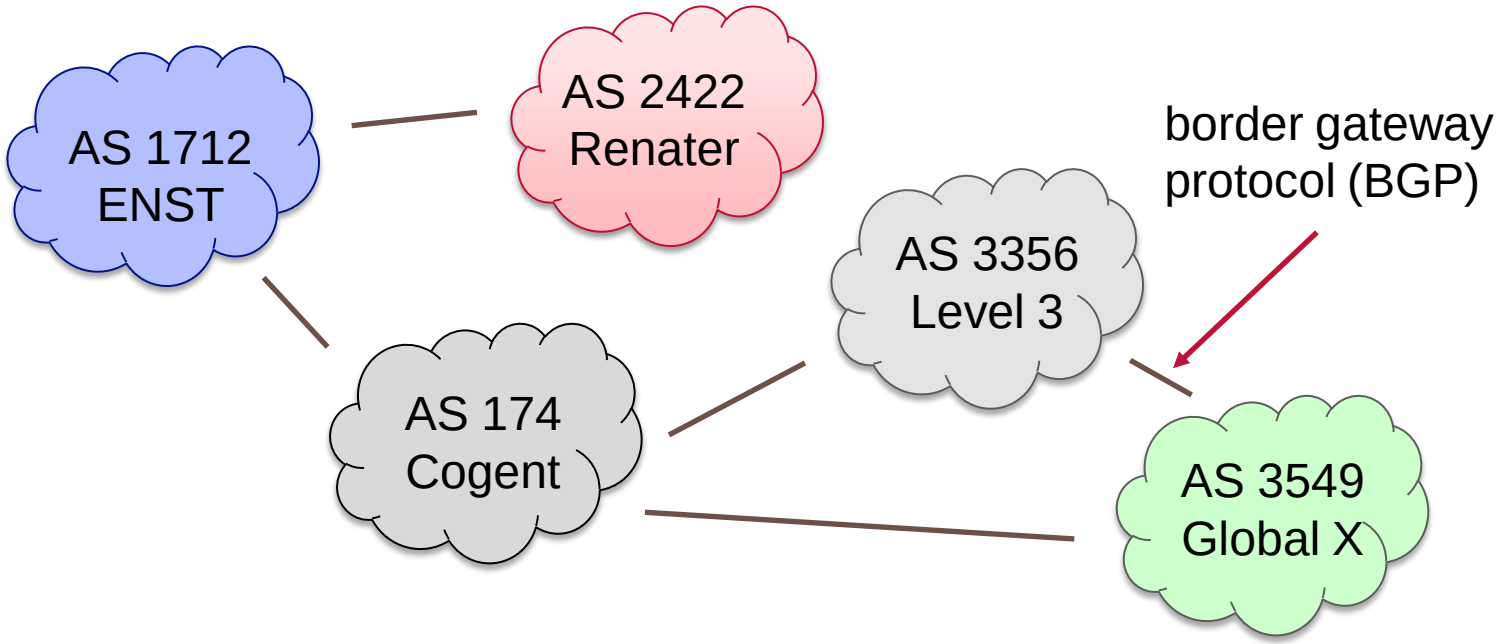
L’organisation de l’Internet

- AS - Autonomous Systems
  - Un AS est un réseau qui est géré par une seule entité administrative



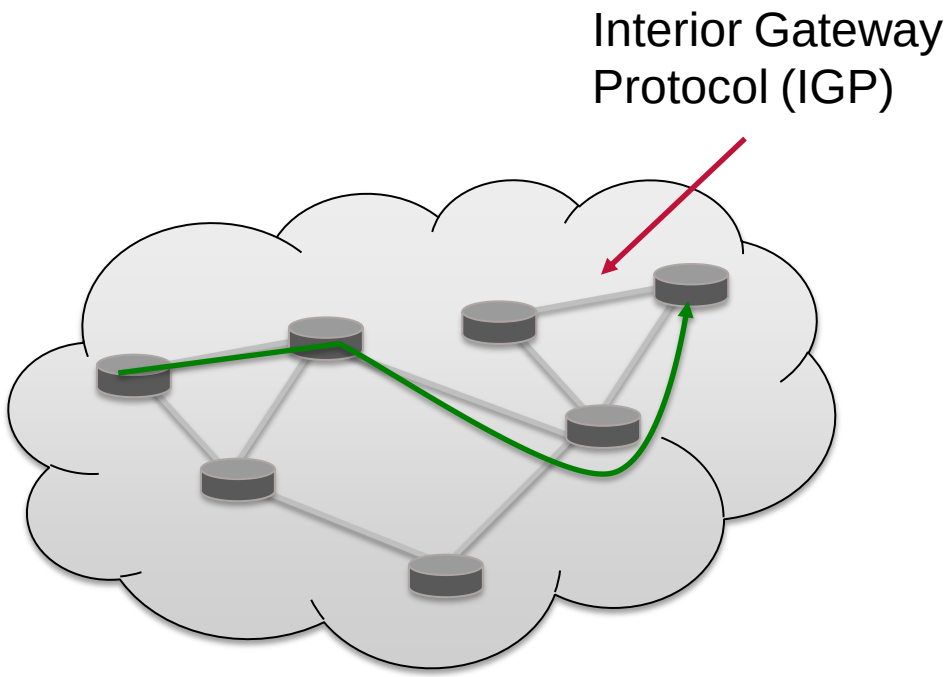
Routage **inter** AS

- Chaque AS:
  - Propage (selectivement) les infos sur le meilleur chemin qui permet de joindre les préfixes administrés par l'AS
  - Attende les infos propagées par les autres ASes

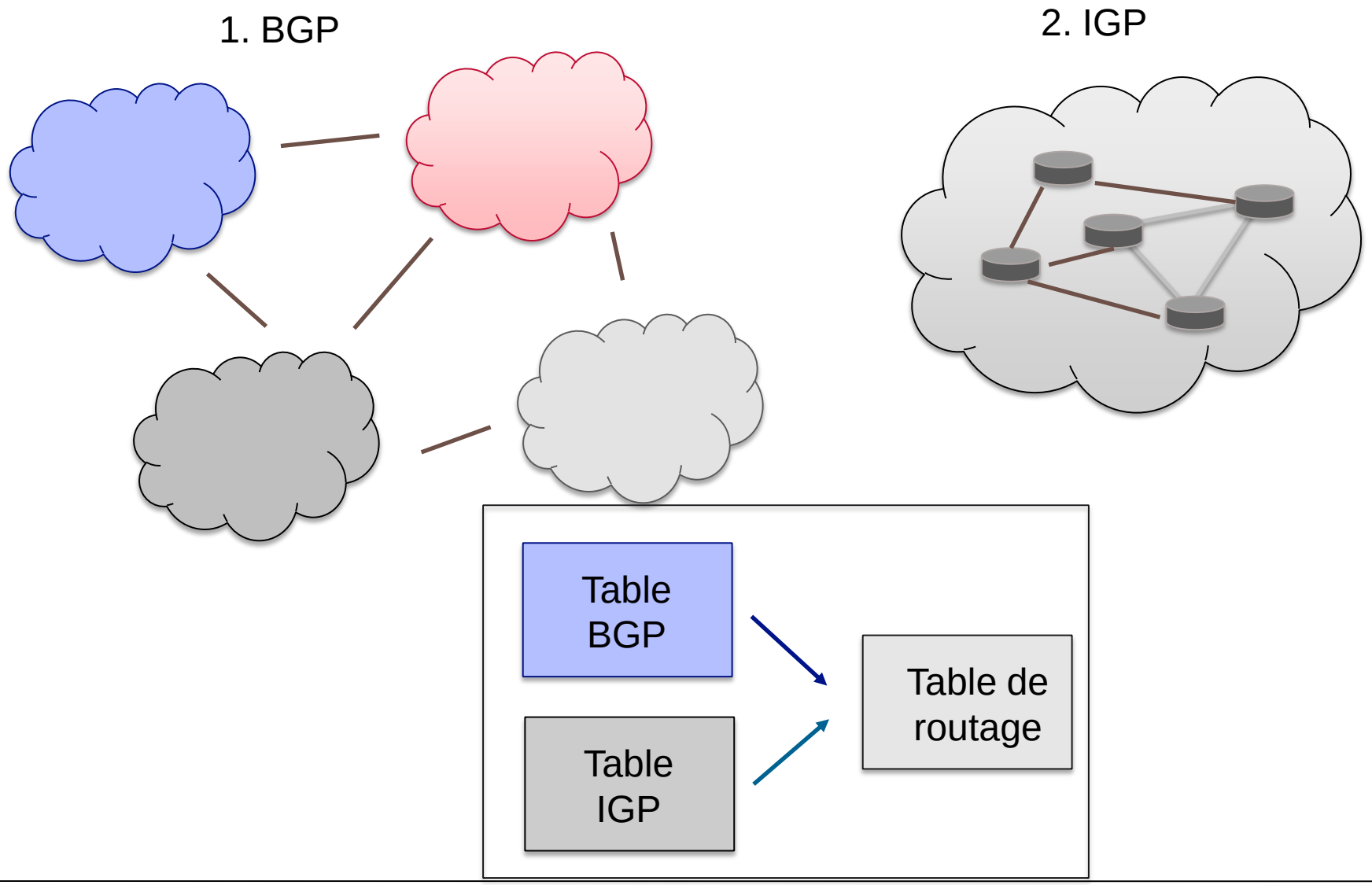


Routage **intra** AS

- Chaque AS:
  - Géstion autonome de son propre routage
  - Routage basé sur le plus-court-chemin



Les deux types de routage



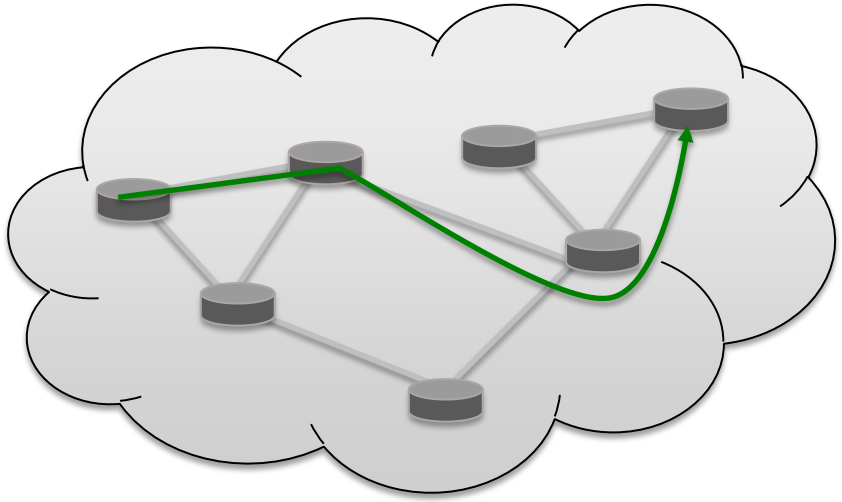
IGP

(Intra-Domain Routing)

Interior Gateway Protocol – IGP

- Principe
  - Choisir le “shortest-path” à l’intérieur d’un AS
  - Basé sur une métrique **additive :**

$$f\left(Link_A + Link_B\right) = f\left(Link_A\right) + f\left(Link_B\right)$$



Les algorithmes de routage

Distance Vector

- I tell you all my “best” routes for all destinations that I know and you tell me yours.
- DV construit une **topologie simplifié** à partir d’une perspective locale
- E.g. **RIP** (Routing Information Protocol)

Link State

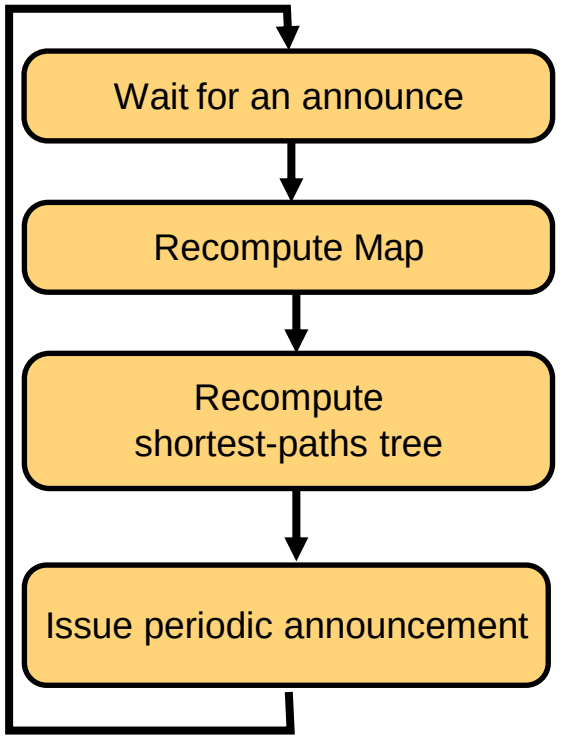
- I announce to everyone about my links and the addresses I originate on each link and listen to everyone’s announcement.
- LS construit une **topologie complète** du réseau
- E.g. OSPF (Open Shortest-Path First)

OSPF

(Open Shortest-Path First)

Link State (OSPF) Formally Defined

- Is an instantiation of the Dijkstra Algorithm:
  - Set: $i = 0, S_0 = \{u_0 = s\}, L(u_0) = 0,$ and $L(v) = \infty,$ for $v \neq u_0$
  - Compute: $\forall v \in (V \setminus S_i)\, L(v) = \min\{L(v), L(u_i) + d_v^{u_i}\}$
  - Select: $u_{i+1} = v' : L(v') = \min_{\forall v \in (V \setminus S_i)}\{L(v)\}$
  - Update: $S_{i+1} = S_i \cup u_{i+1}, i = i + 1$
  - If $i =\mid V \mid -1$ Stop, otherwise go to 2



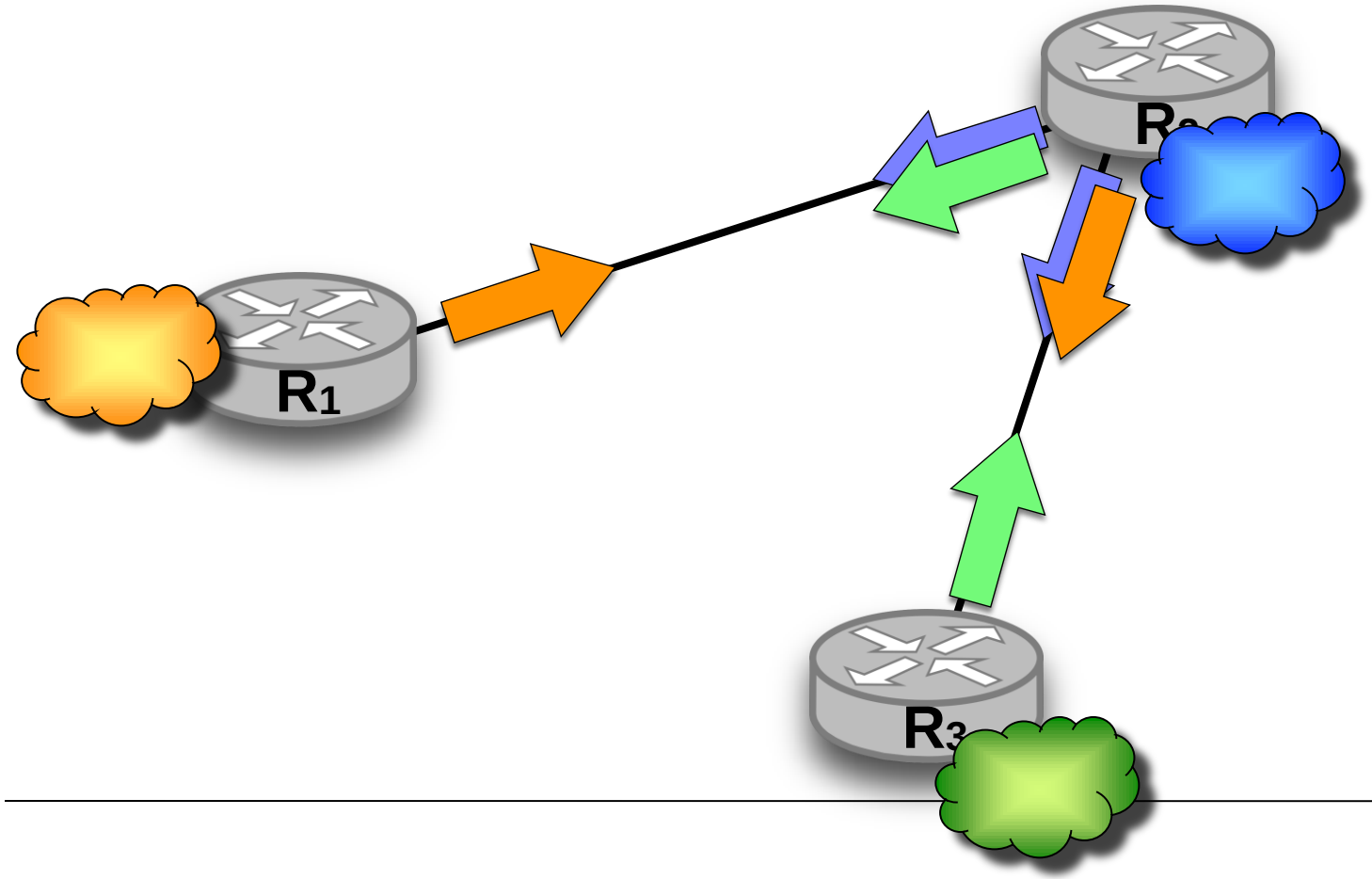
OSPF: Open Shortest-Path First Informally Defined

- Le routeur propage les infos sur toutes ses connexions (liens), et l'état des liens (up/down)
  → les infos propagées s'appellent **announcements**
- Le routeur propage aussi les adresses qu'il garde sur chaque lien
- Le routeur reste à l'écoute pour les announcements des autres routeurs
- À partir de ça, il est possible de construire une topologie pour chaque lien (map)
- Gràce à la map du réseau, le routeur peut calculer le shortest path pour tous préfixes

Assumption : chaque routeur a construit la topologie du réseau et on a « consensus » que tout le monde partage la meme vision de la topologie (et le même choix des plus courts chemins)

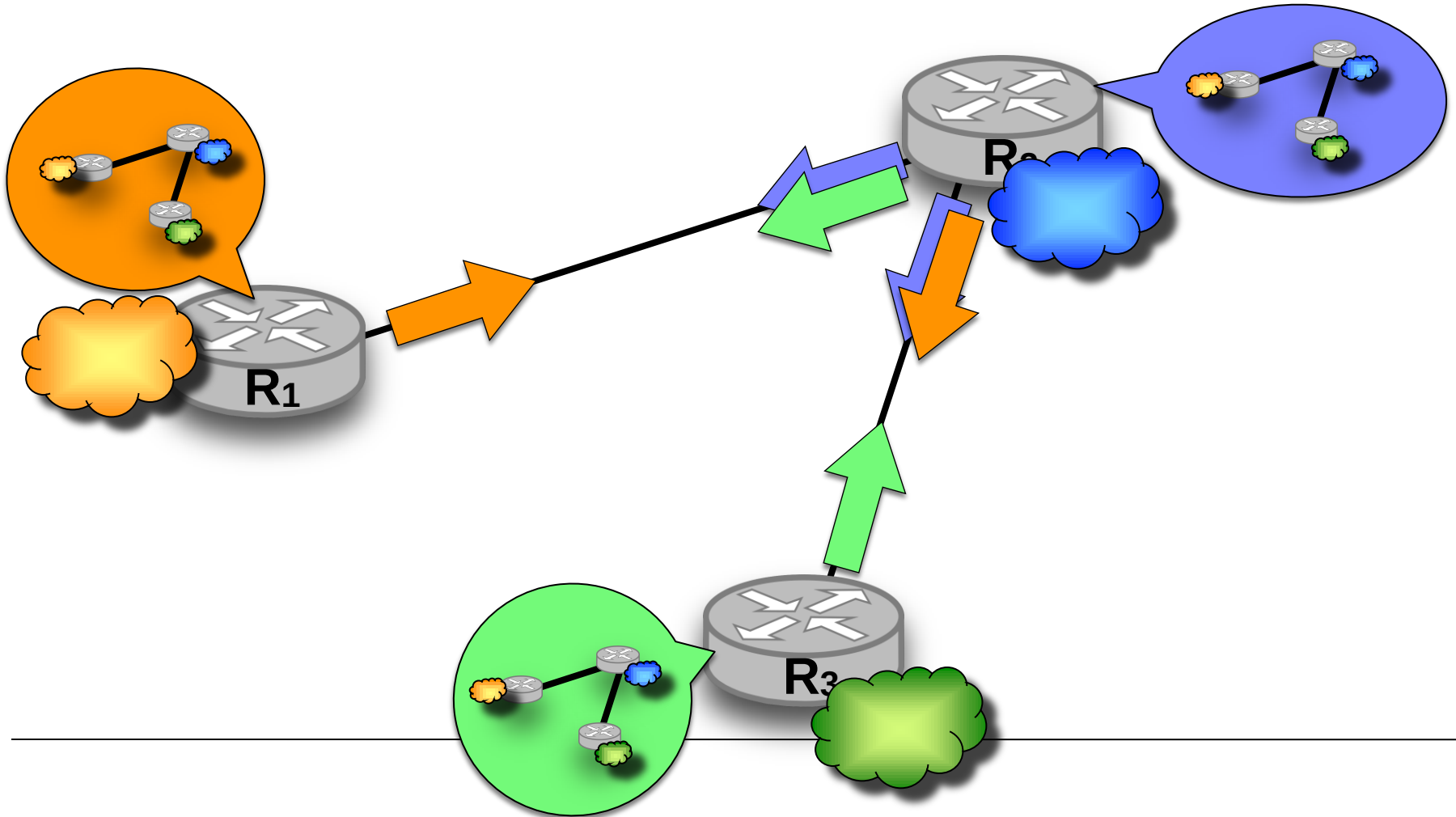
OSPF: Link-State Advertisements

- Routing information (reachability, link state) is broadcasted



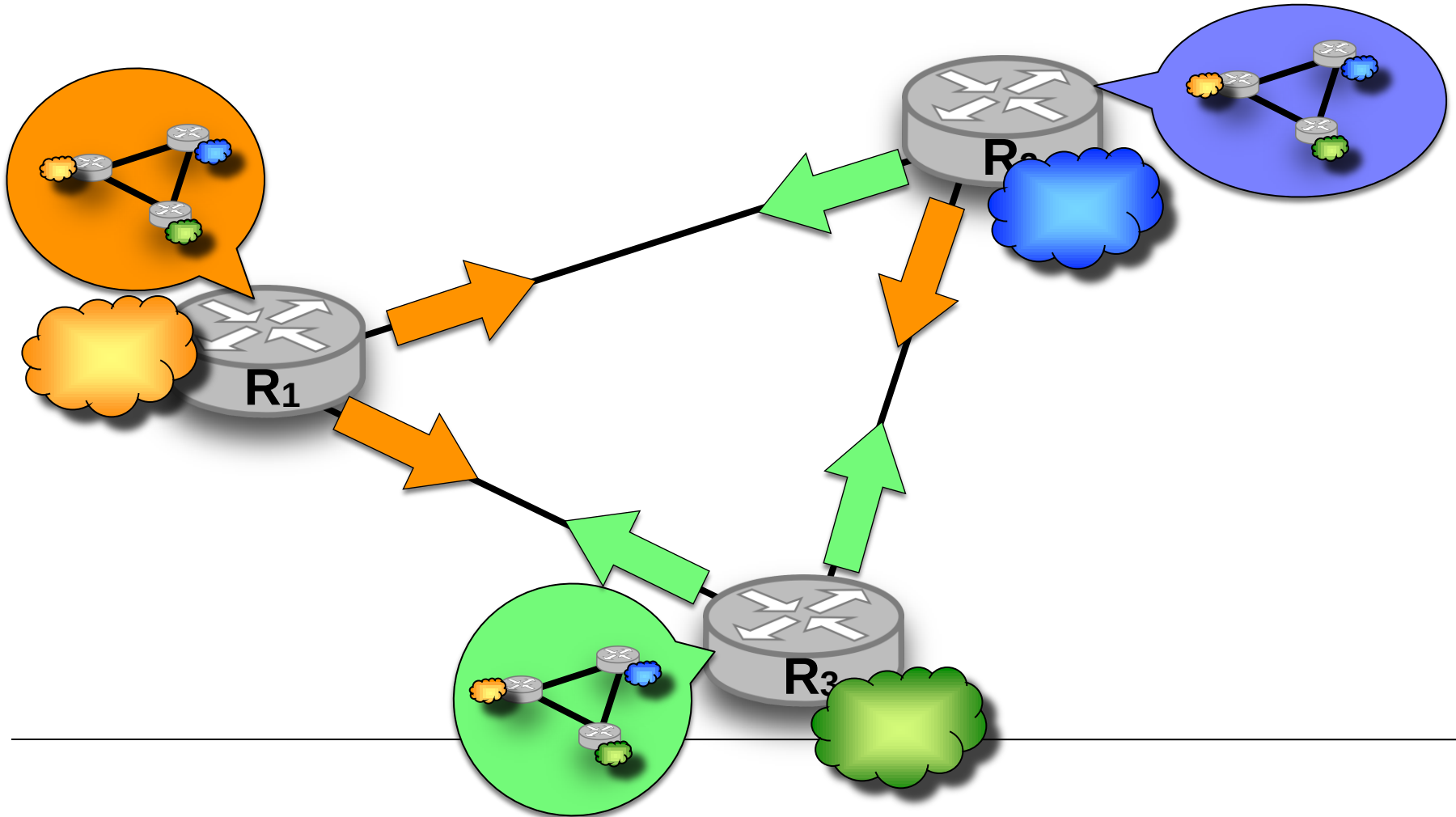
OSPF: Link-State Advertisements

- Routers build global view of the topology
- Routing table obtained by computing the shortest path on the topology



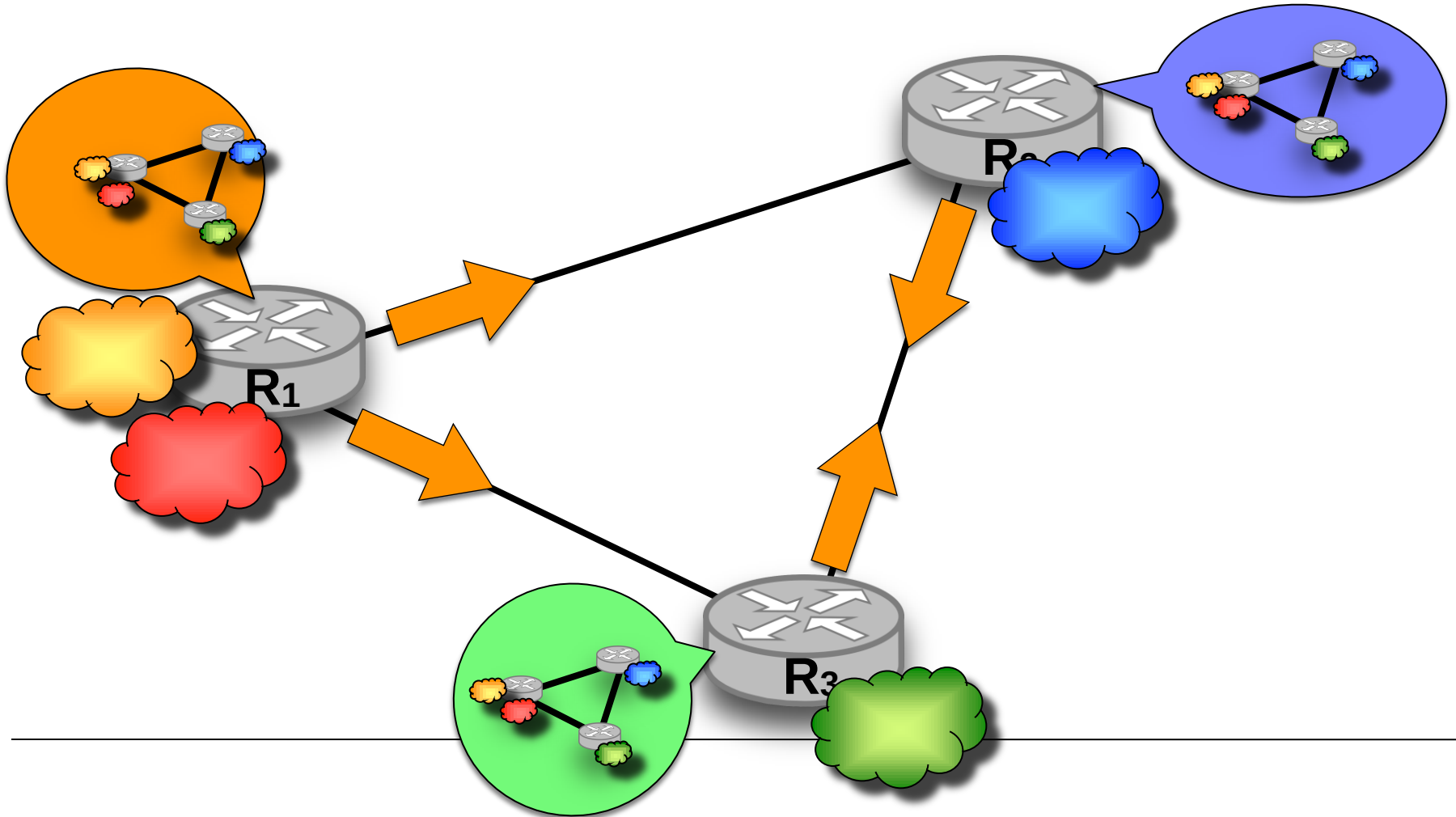
OSPF: Link-State Advertisements

- Convergence is rapid
- One broadcast round and everybody has the same view



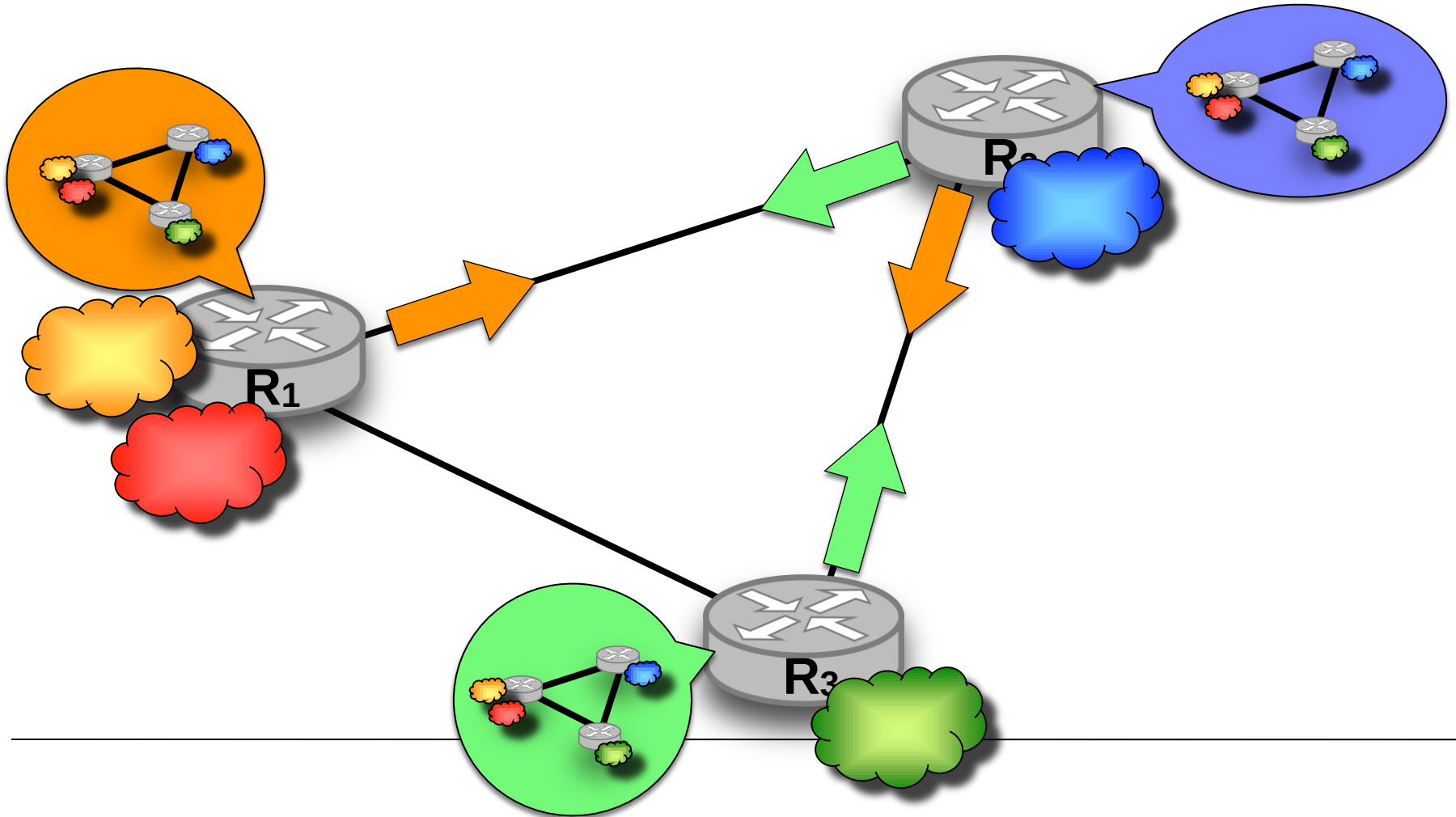
OSPF: Link-State Advertisements

- Convergence is rapid
- One broadcast round and everybody has the same view



OSPF: Link-State Advertisements

- Convergence is rapid even in case of failures
- One broadcast round and everybody has the same view



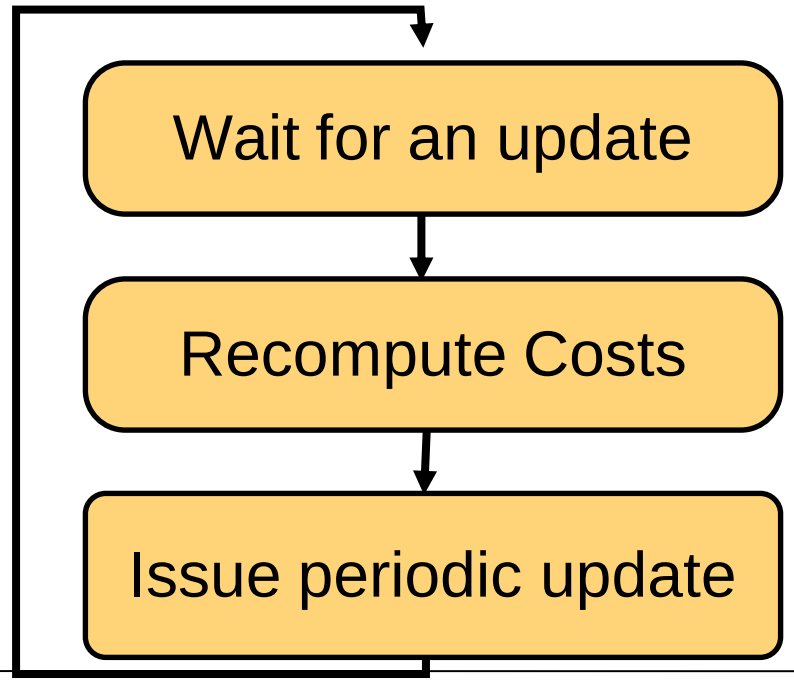
RIP

(Routing Information Protocol)

Distance Vector (RIP) Formally Defined

- Algorithme de **Bellman-Ford**
  - Define Dx(Y) := cost of the least-cost path from X to Y
  - Then:

$$d_{(me)}(Dst) = \min_{\binom{All\ \ my}{neighbors}} \{d_{(me)}(n_x) + d_{(n_x)}(Dst)\}$$

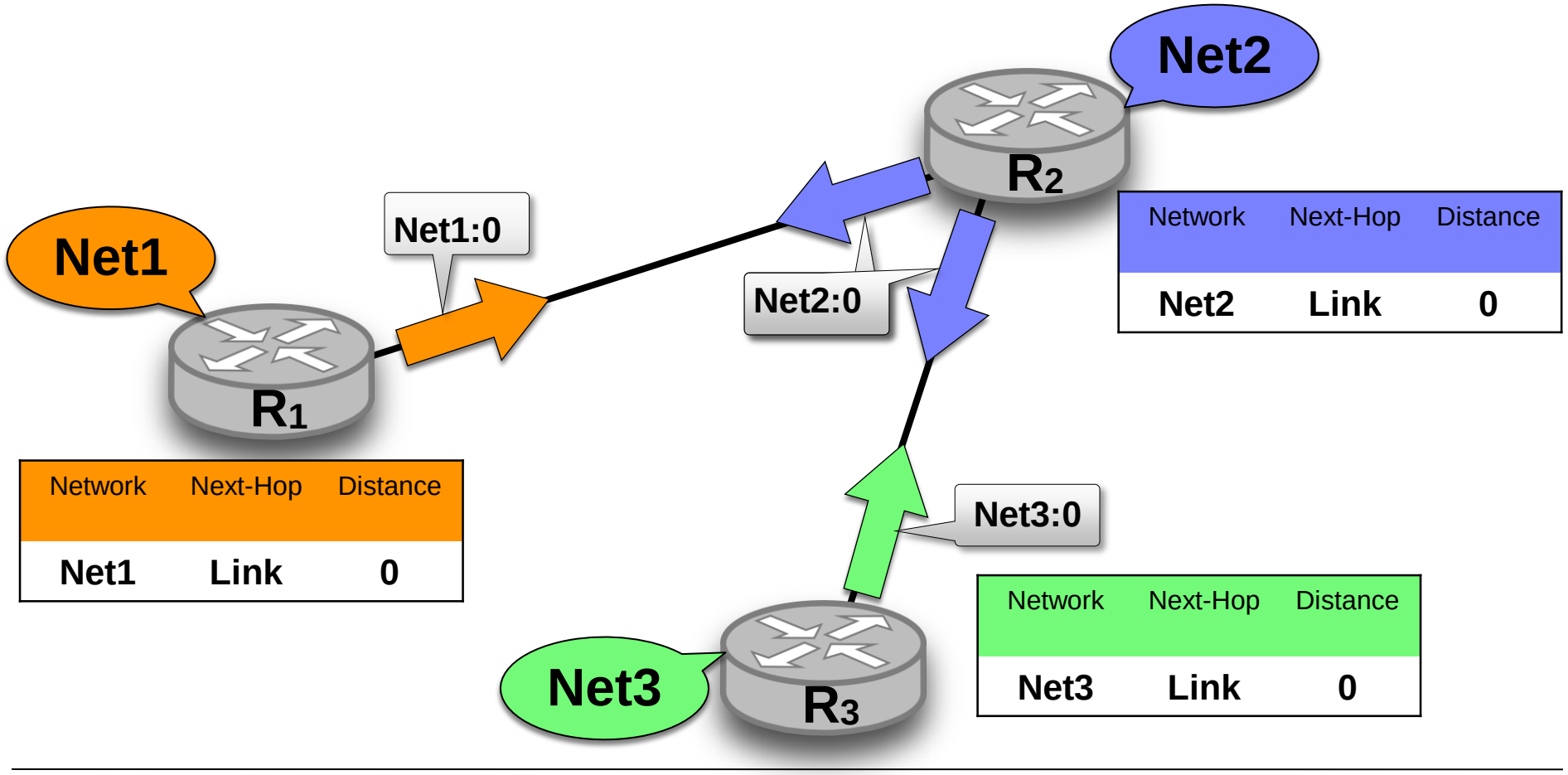


RIP: Routing Information Protocol Informally Defined

- Le routeur propage les “best” chemins pour toutes destinations qui sont connues
- À partir de cette connaissance (locale) il construit une topologie simplifiée
- Si jamais d’autres chemins sont meilleurs, le routeur va mettre à jour avec ces informations
- Dans le cas d’une mise à jour, le routeur envoie des advertisements.

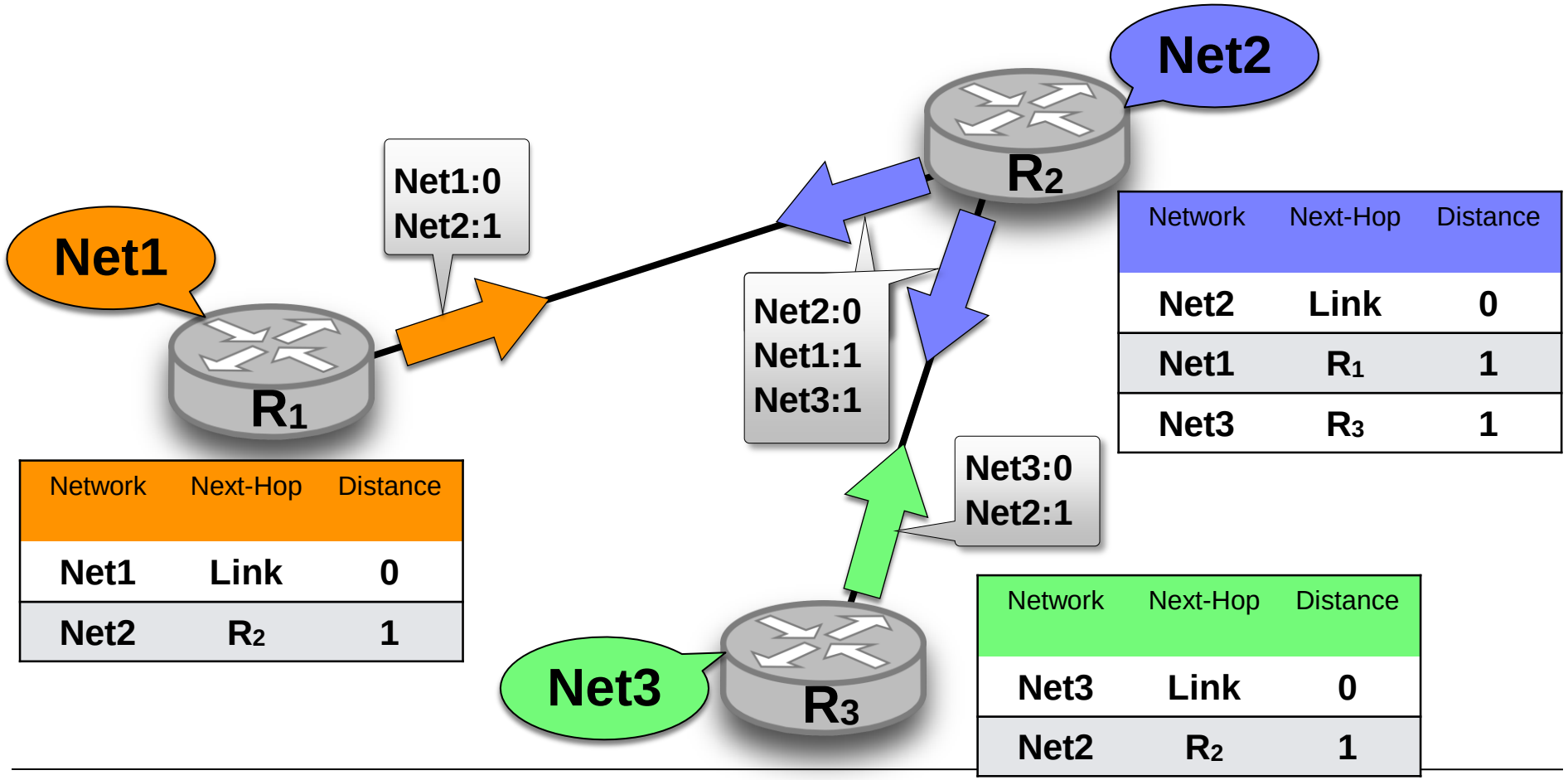
RIP: Initialisation

- Chaque routeur connaît les réseaux directement connectés
- Les tables de routage sont envoyés seulement aux voisins



RIP: Transient Phase

- (si necessaire) les tables de routage sont mis à jour dans le routeur
- Les tables à jour sont envoyées (seulement aux voisins)



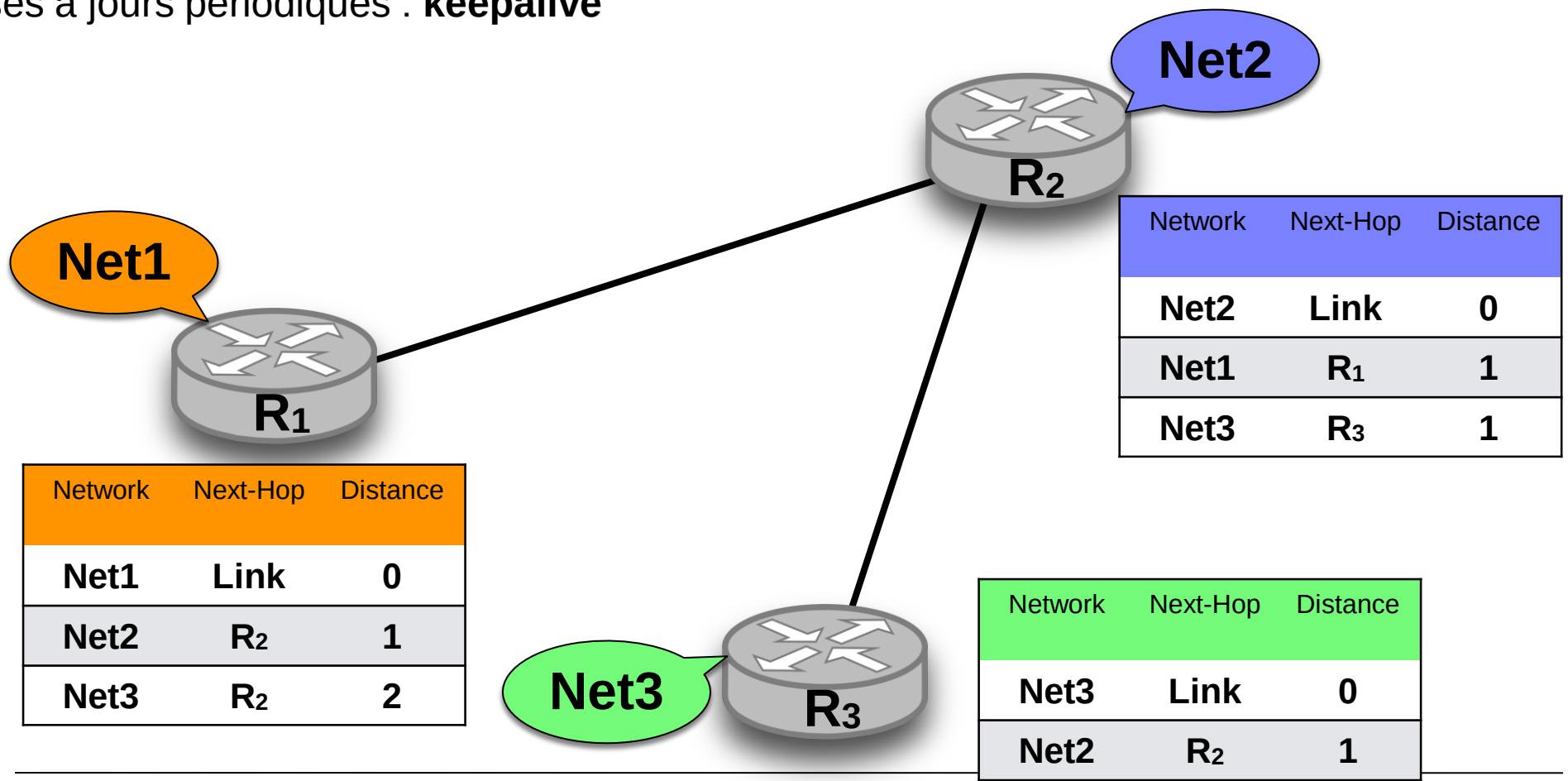
Network	Next-Hop	Distance
Net1	Link	0
Net2	R_2	1

Network	Next-Hop	Distance
Net2	Link	0
Net1	R_1	1
Net3	R_3	1

Network	Next-Hop	Distance
Net3	Link	0
Net2	R_2	1

RIP: Convergence

- Convergence : tout le monde a atteint les mêmes informations de routage
- Régime Permanent (sauf si changement de topologie)
  - Mises à jours periodiques : **keepalive**



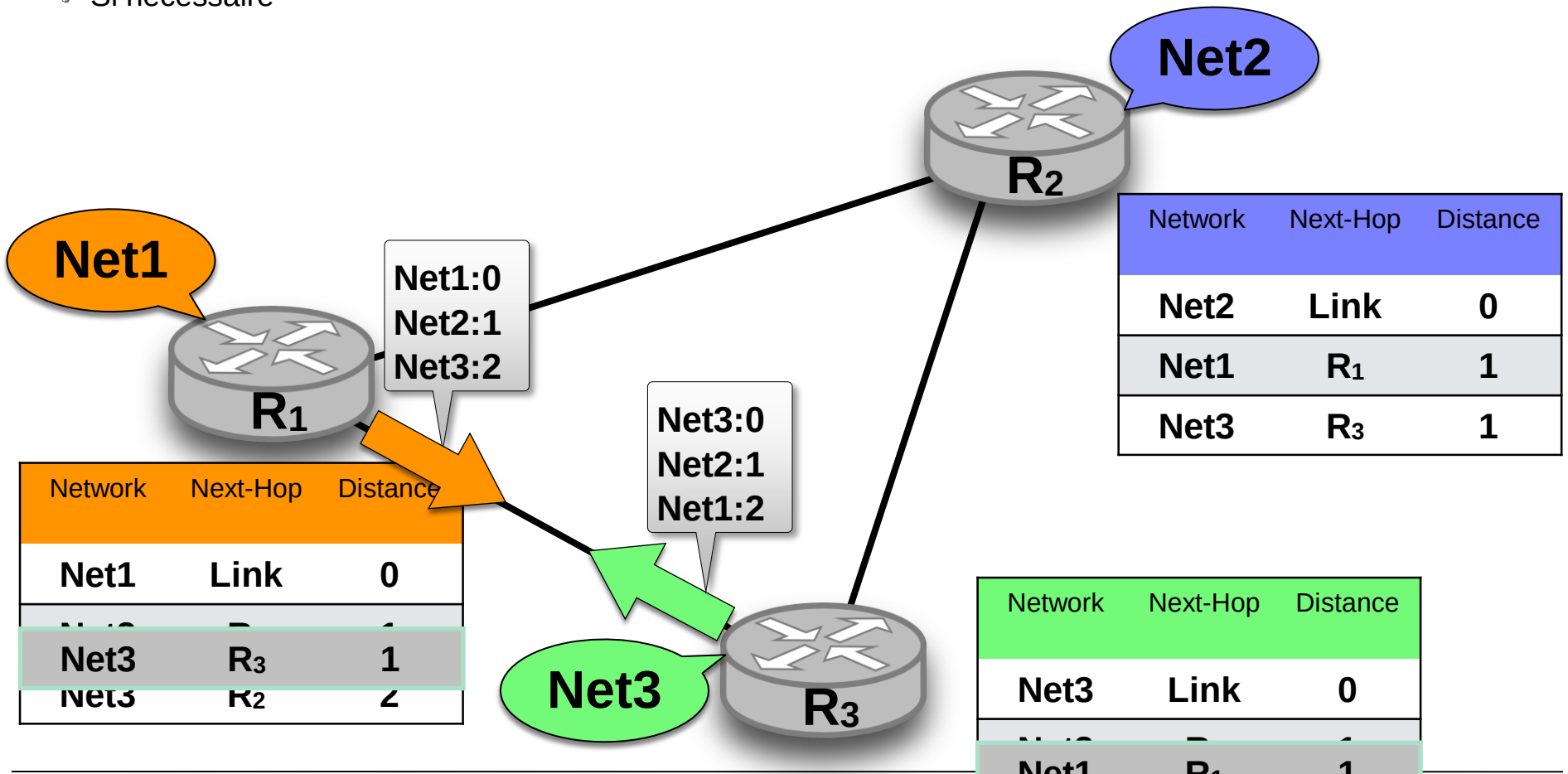
Network	Next-Hop	Distance
Net1	Link	0
Net2	R_2	1
Net3	R_2	2

Network	Next-Hop	Distance
Net2	Link	0
Net1	R_1	1
Net3	R_3	1

Network	Next-Hop	Distance
Net3	Link	0
Net2	R_2	1
Net1	R_2	2

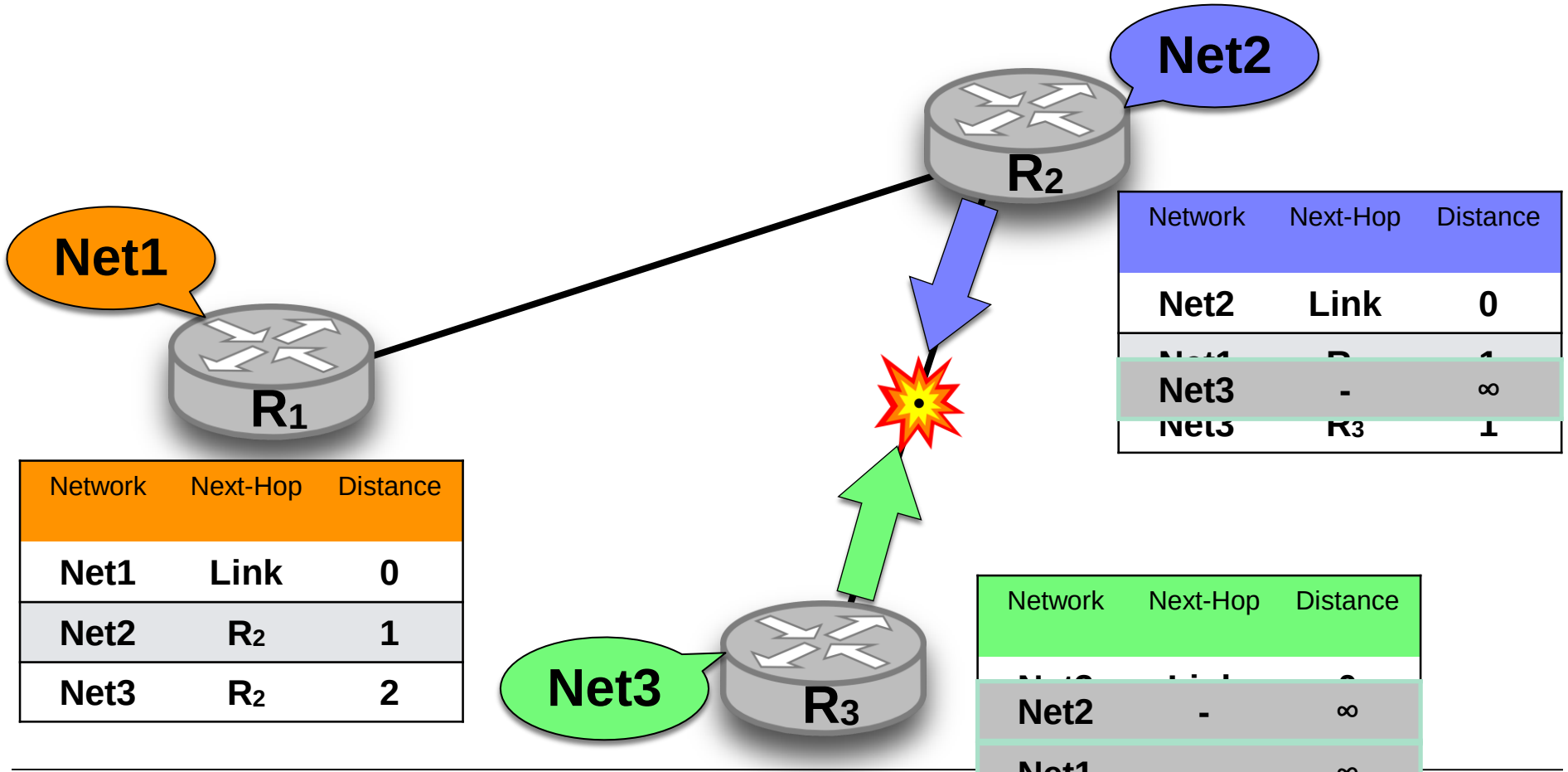
RIP: Adaptation

- E.g., ajoute d'un lien : Distance Vector (DV) est envoyé dans le nouveau lien
- Ceci permet de mettre à jour les tables de routage
  - Si nécessaire



RIP: Failure Detection

- Basée sur des **timeouts**
  - Les updates ne sont pas reçus pour un certain temps



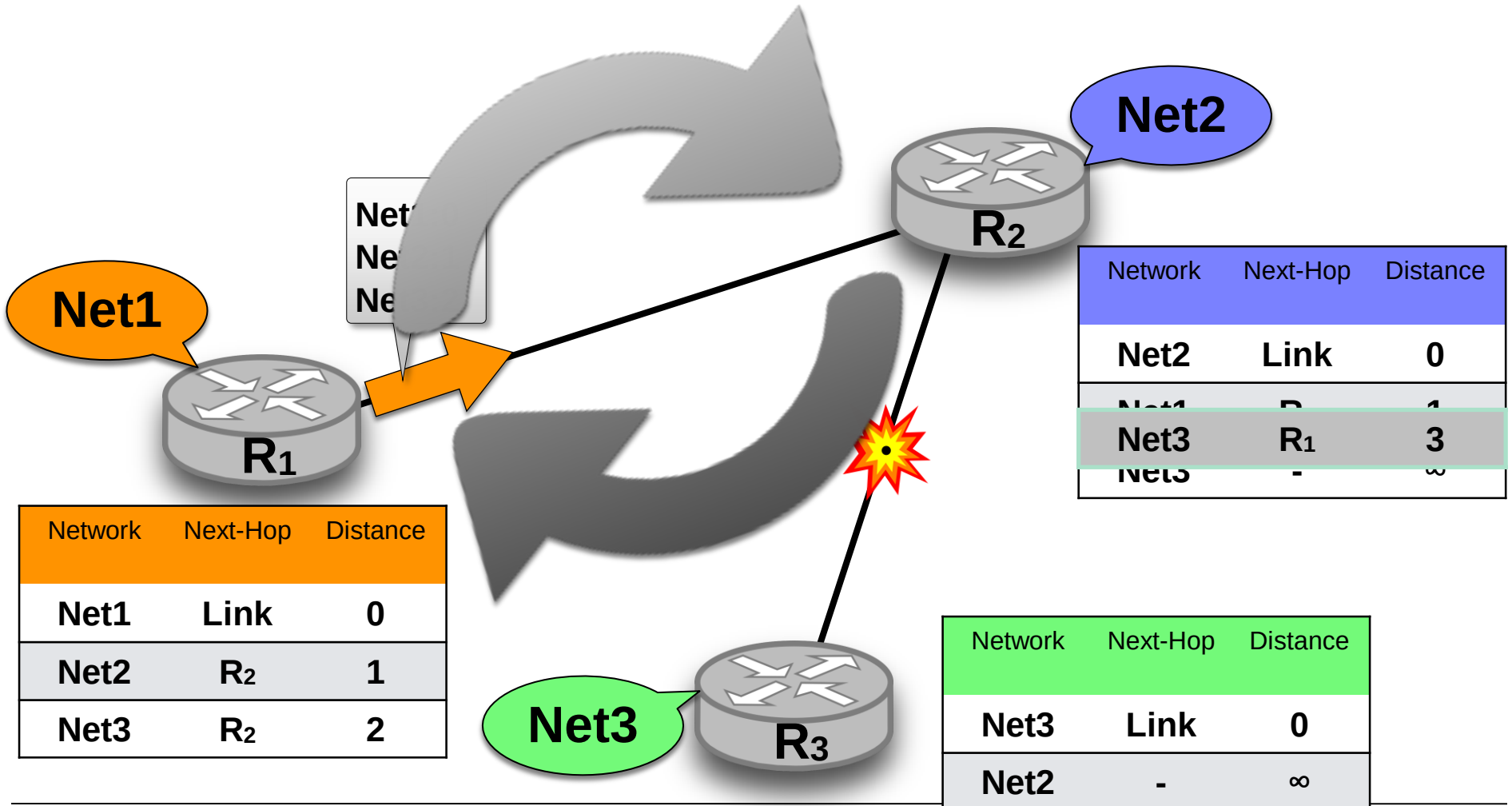
Network	Next-Hop	Distance
Net1	Link	0
Net2	R_2	1
Net3	R_2	2

Network	Next-Hop	Distance
Net2	Link	0
Net3	-	∞
Net3	R_3	1

Network	Next-Hop	Distance
Net2	-	∞
Net1	-	∞
Net1	R_2	2

RIP: Loops Happens

- Vision partielle : peut générer des boucles



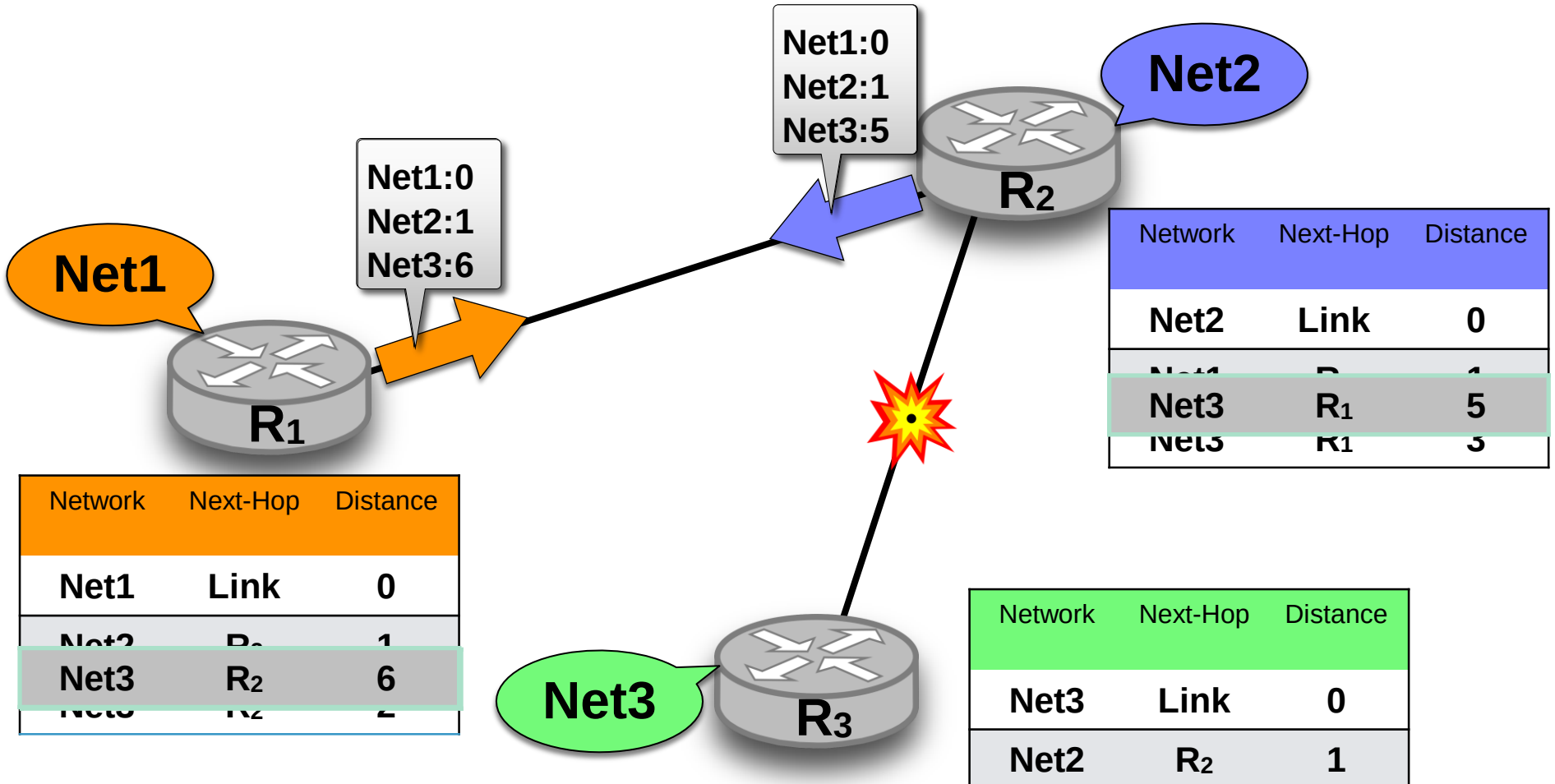
Network	Next-Hop	Distance
Net2	Link	0
Net1	R1	1
Net3	R1	3
Net3	-	∞

Network	Next-Hop	Distance
Net1	Link	0
Net2	R2	1
Net3	R2	2

Network	Next-Hop	Distance
Net3	Link	0
Net2	-	∞
Net1	-	∞

RIP: **Count to infinity**

- Les deux routeurs font une escalade des coûts jusqu'à l'infini
- In original RIP : Infinity = 16
  - Updates chaque 30 secs → 8 minutes de boucle...



Network	Next-Hop	Distance
Net2	Link	0
Net3	R_1	5

Network	Next-Hop	Distance
Net1	Link	0
Net3	R_2	6

Network	Next-Hop	Distance
Net3	Link	0
Net2	R_2	1
Net1	R_2	2

RIP vs OSPF

RIP

- Très simple
- Très lent pendent convergence
  - Good news : propagation plutot rapide
  - Bad news : propagation très lente
- Difficilè détecter des boucles
- Flat
- Non scalable (16 hops = infinity)

OSPF

- Plutôt complexe
  - Neighbour adjacency*
  - Topology Database
  - Shortest Path Routing Table

- Très efficace et à convergence rapide
- Les boucles et/ou inconsistances sont faciles à détecter
- Hiérarchique*
- Scalable

* En détail dans GIN201

Link State vs. Distance Vector

Criterion	Distance-Vector	Link-State
Complexity (CPU)	Simple	Complex
Memory Usage	Low	High
Load	High	Low (*)
Convergence	Slow**	Fast

(*) Low refresh period (30minutes), but requires broadcasts.

(**) Unless triggered updates used, but still slower than link-state

Done for today

You are now a Internet Routing Pro!
– Plus de détail seront traité dans GIN201 !