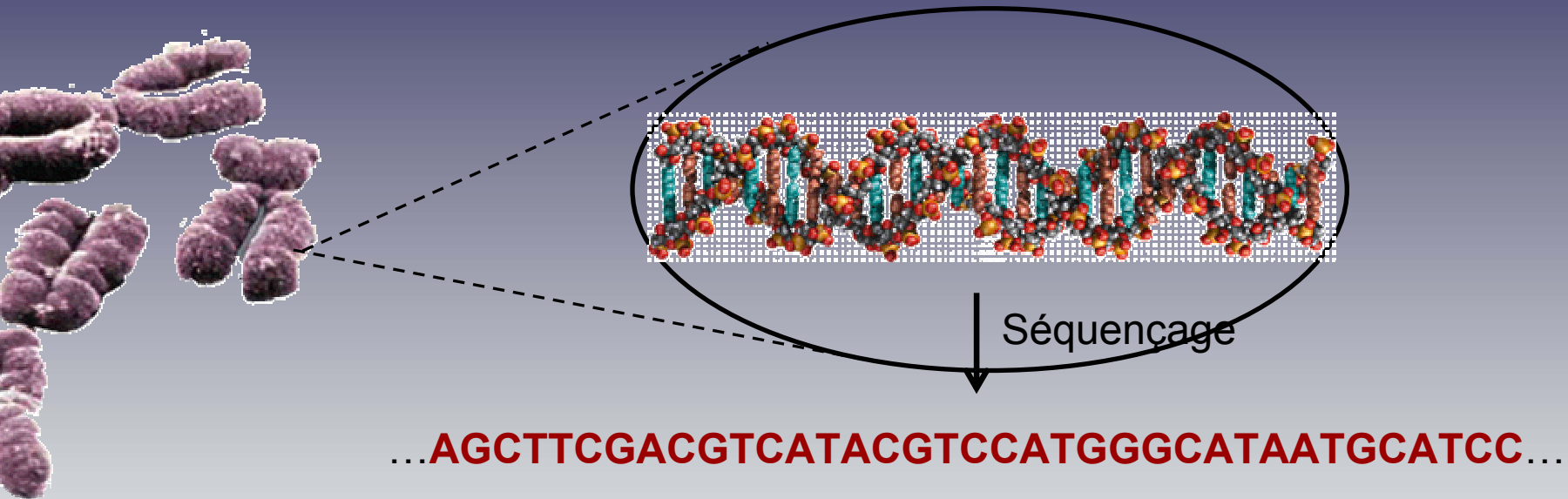


Développement d'une architecture de type GRID pour l'analyse des données d'expression biologiques

O. Delemar, T. Vermat, F. Paillier & M. Deleeuw
GENOME express

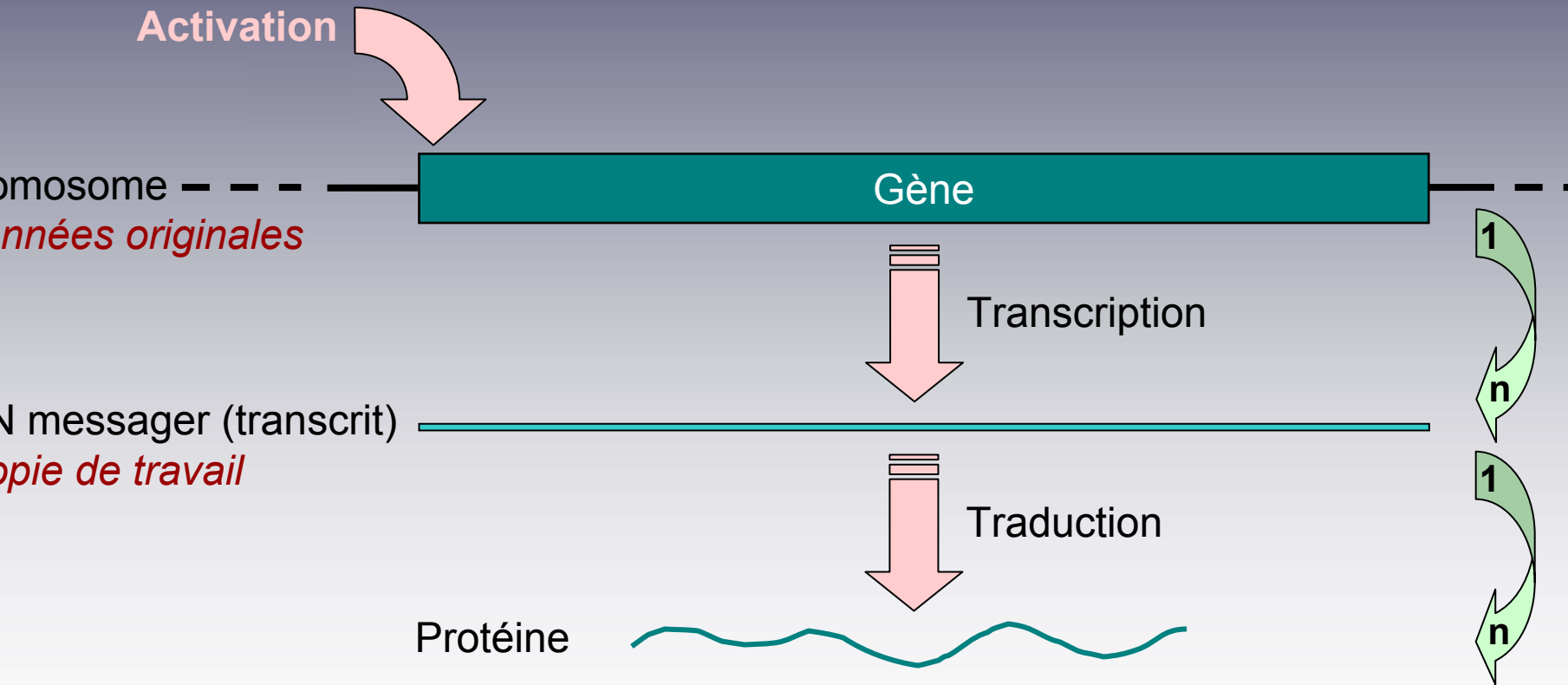
Contexte biologique : Génome et ADN...



- Les chromosomes : support de l'information génétique
- L'ADN : molécule linéaire formée de 4 bases :
l'Adénine, la Cytosine, la Guanine et la Thymine
- On représente l'ADN par une chaîne dans un alphabet de 4 lettres : A, C, G et T

Chez l'homme : 23 paires de chromosomes pour 3 milliards de bases

Contexte biologique : Du gène à la protéine



Contexte biologique : expression des gènes



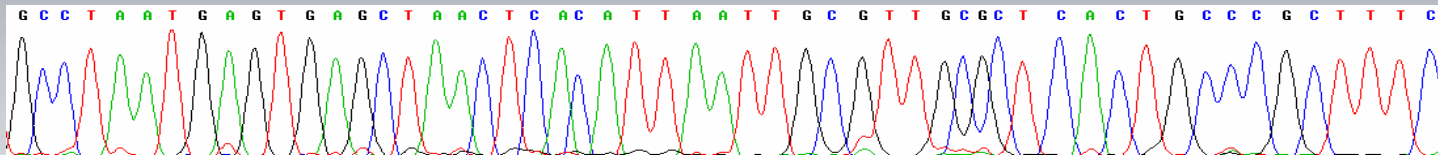
	État 1	État 2
A	x1	x1
B	x2	x6
C	x3	x3
D	x6	x6

Problème : Identifier un grand nombre de transcrits
Chez l'homme, plusieurs milliers de gènes exprimés
étude de 15000 transcrits

Discriminative Analysis of Clone Signatures

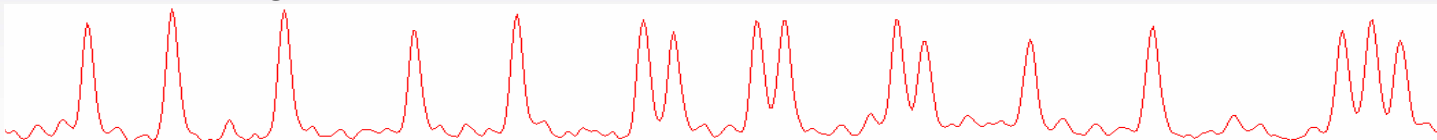
Séquençage d'ADN :

- connaître la succession exacte des bases
- association d'une molécule fluorescente à chaque base
- 1 tube = 1 séquence



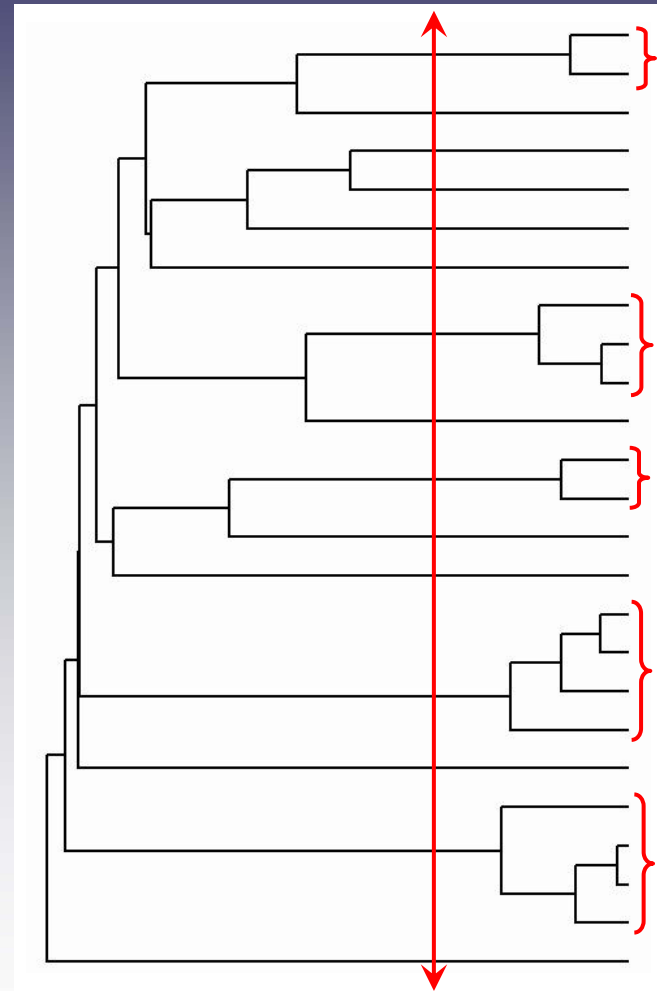
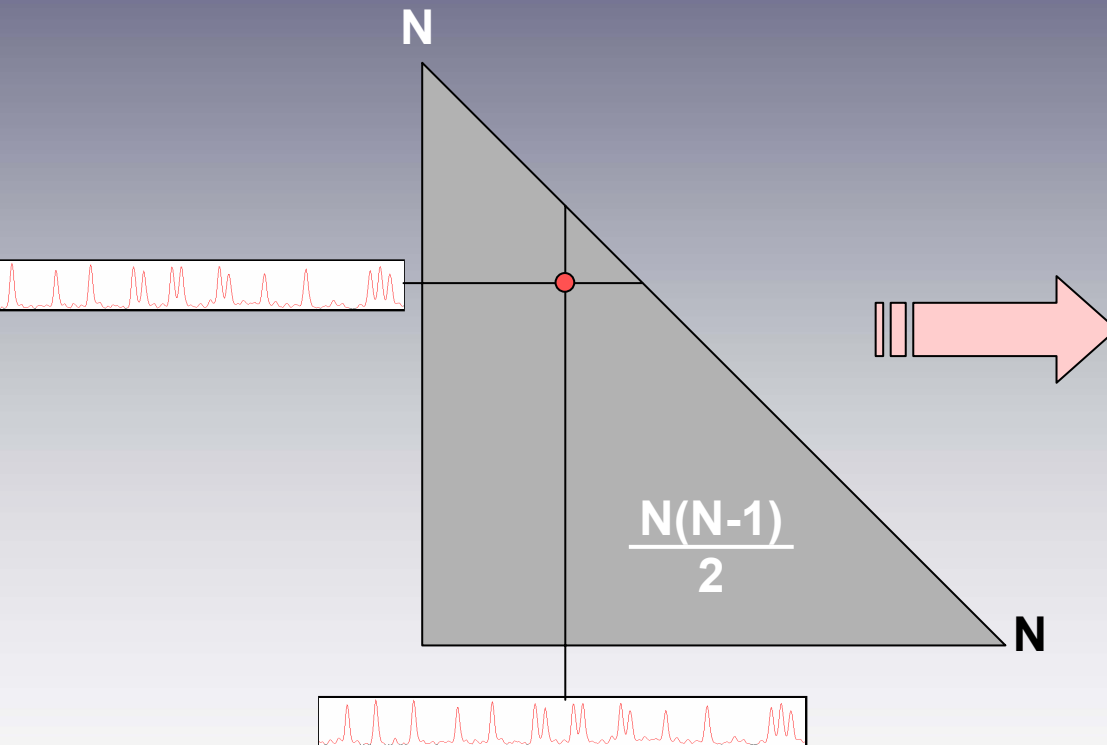
DACS

- identifier un transcrit
- séquençage d'une seule base
- 1 tube = 4 signatures



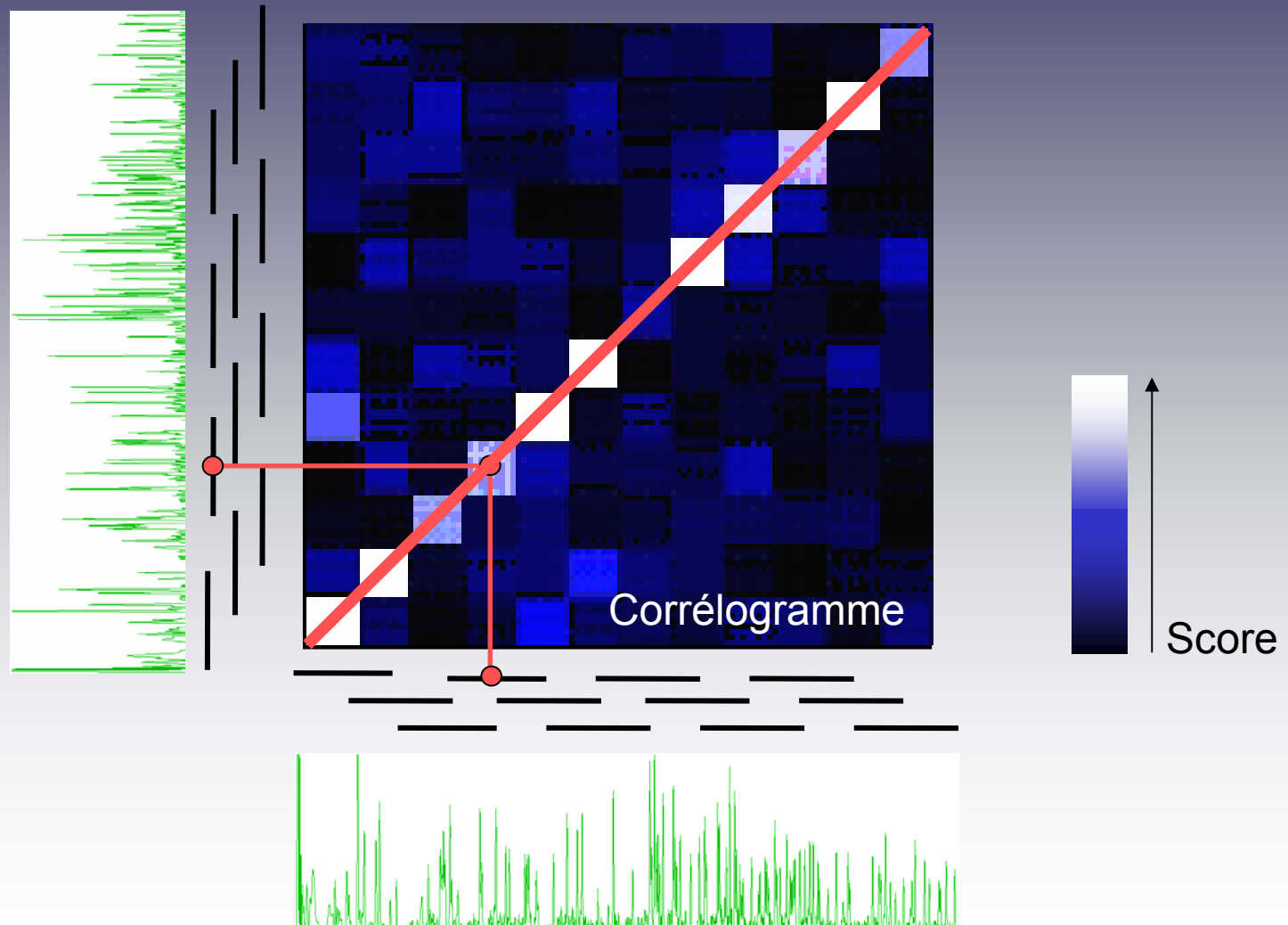
analyse d'expression

Matrice de distances

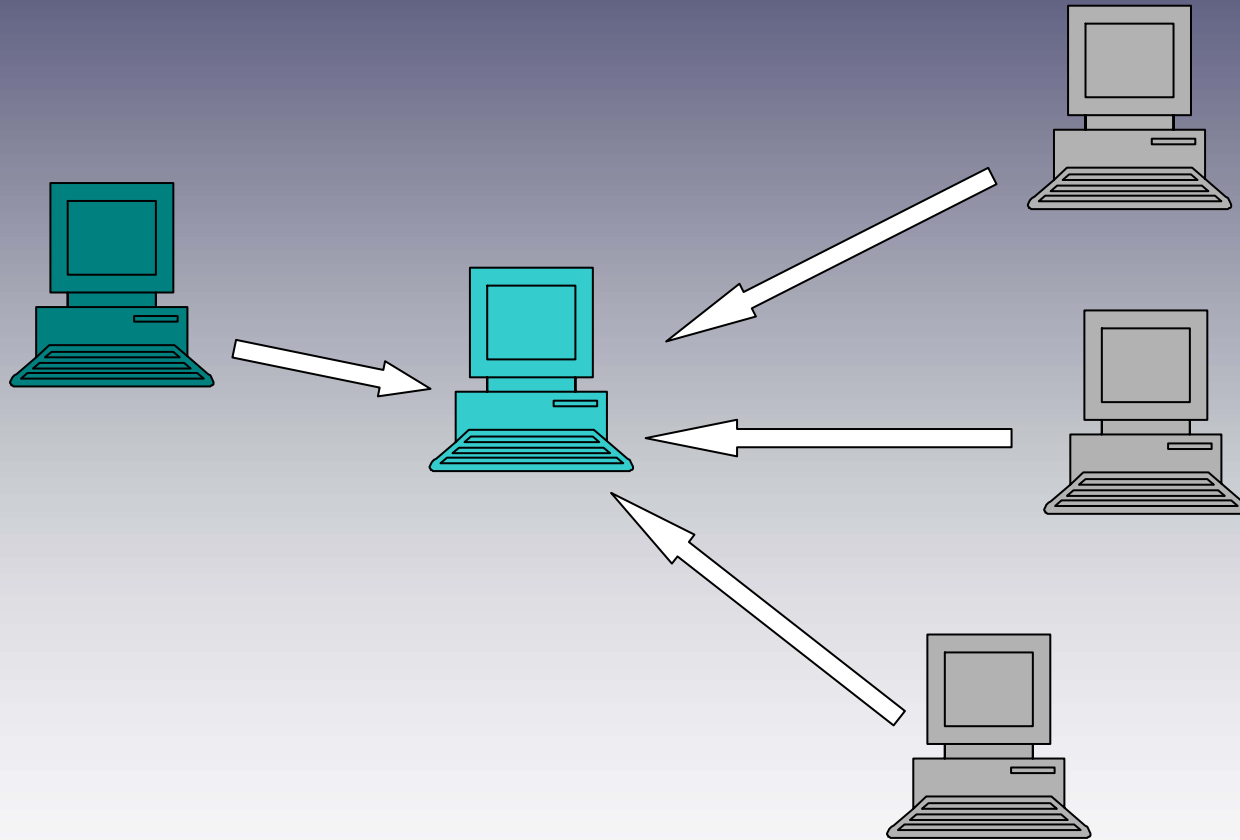


15000 signatures $\Rightarrow$ 49 995 000 comparaisons

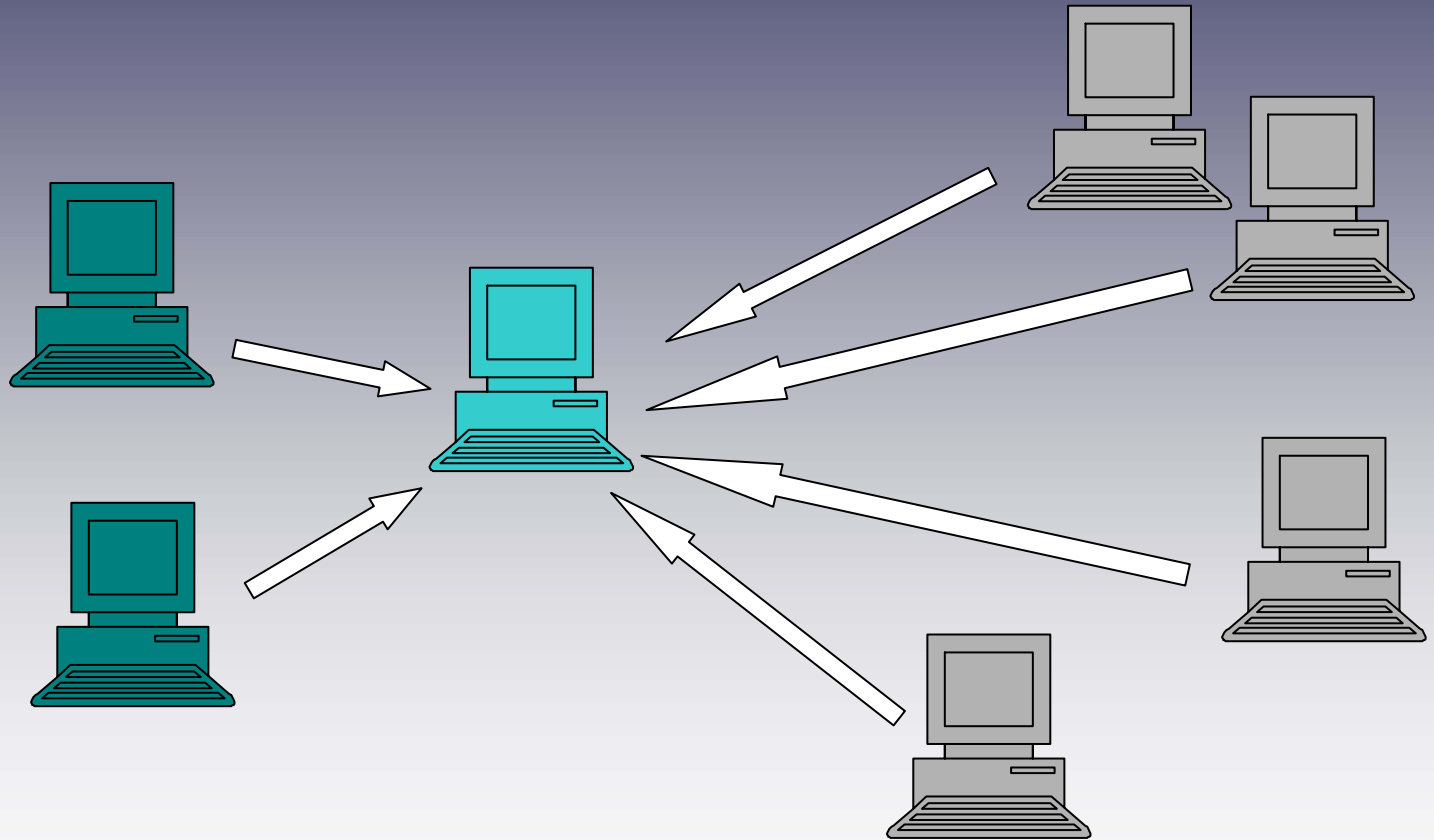
comparaison de signatures



ne architecture de type GRID



ne architecture de type GRID



hoix technologique

Multi plateforme

Traitement de signal

LabVIEW

Prototypage rapide

Communications

schéma de communication

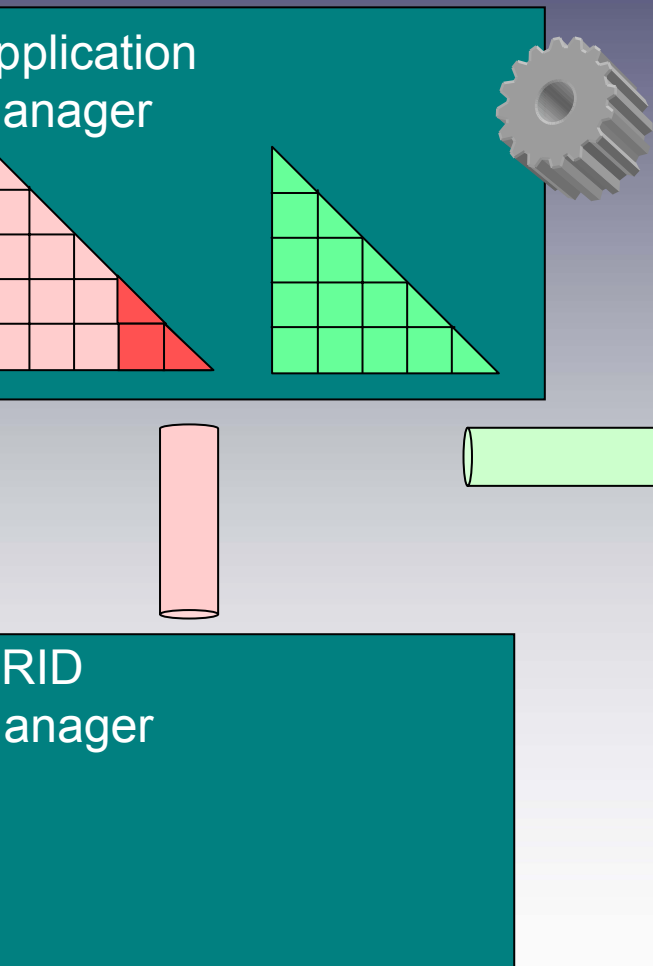


schéma de communication

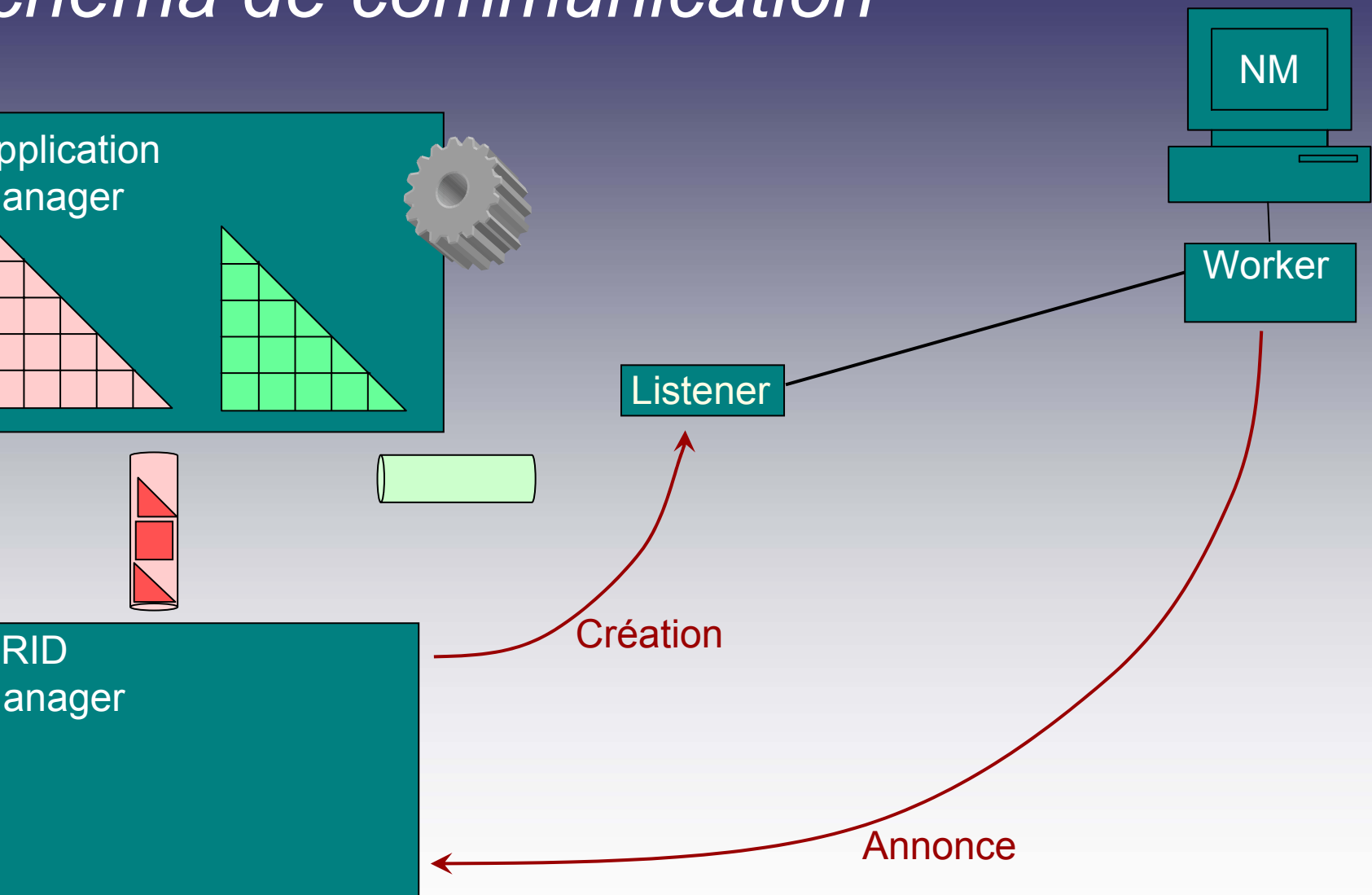


schéma de communication

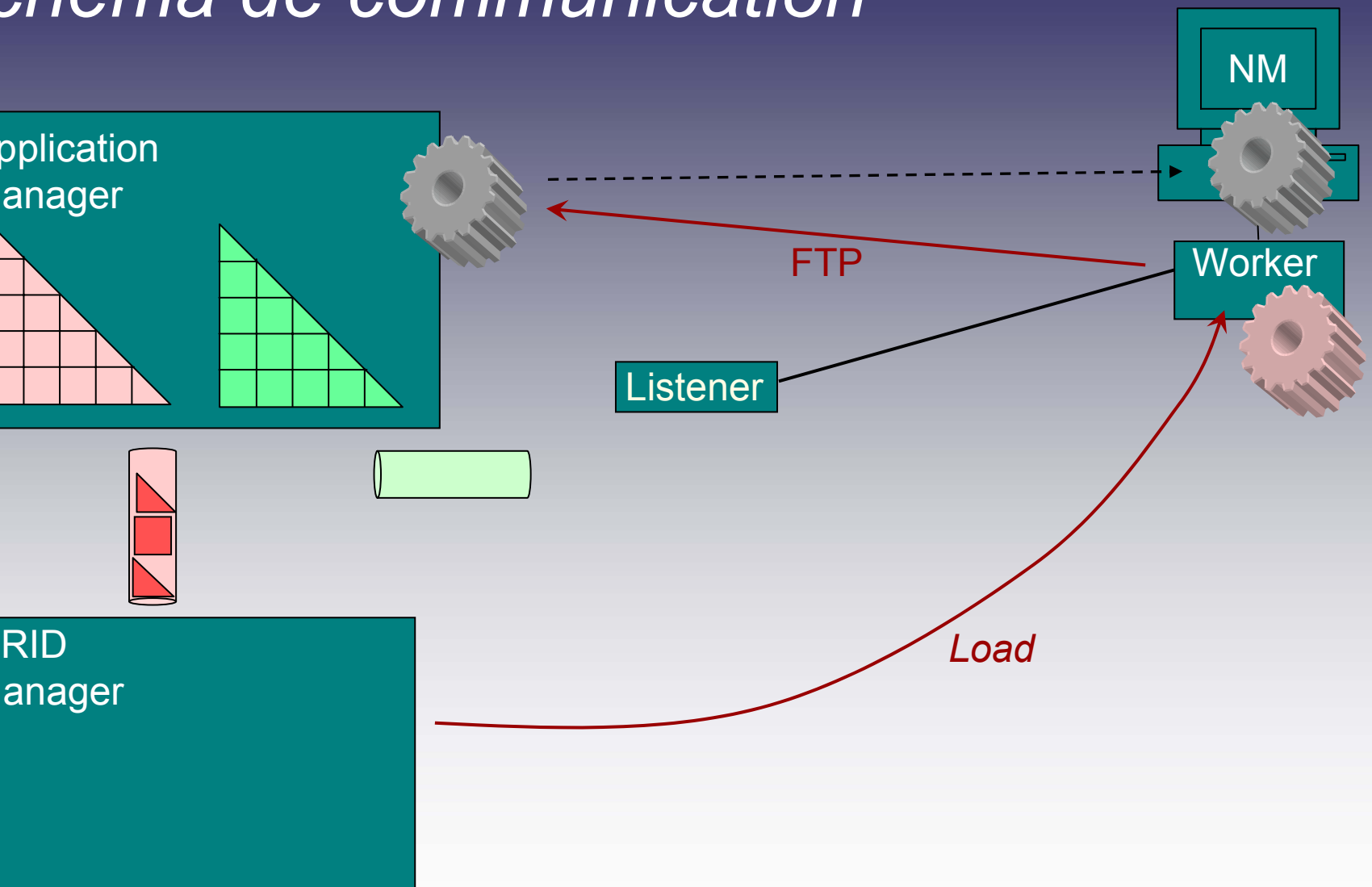


schéma de communication

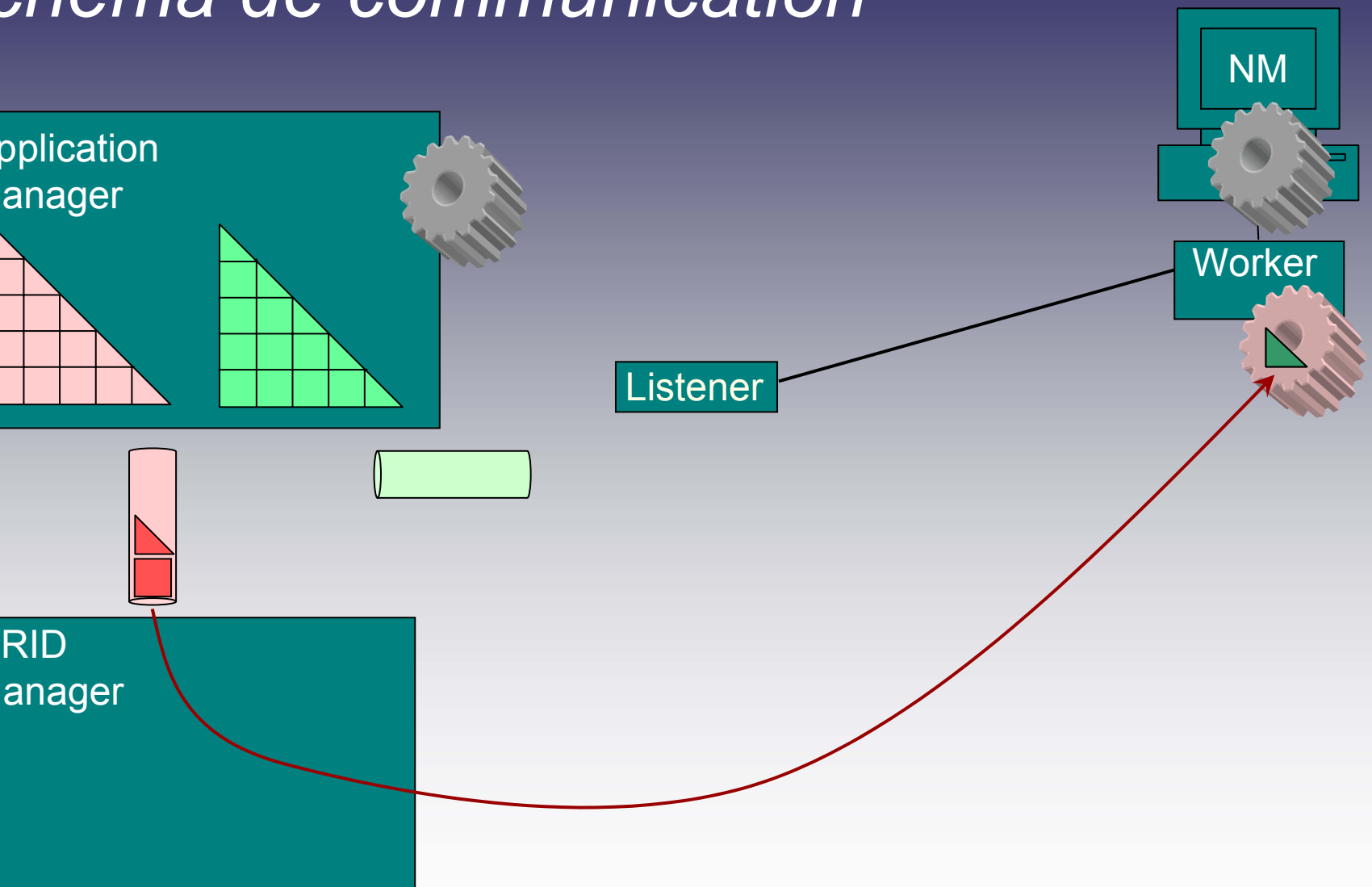
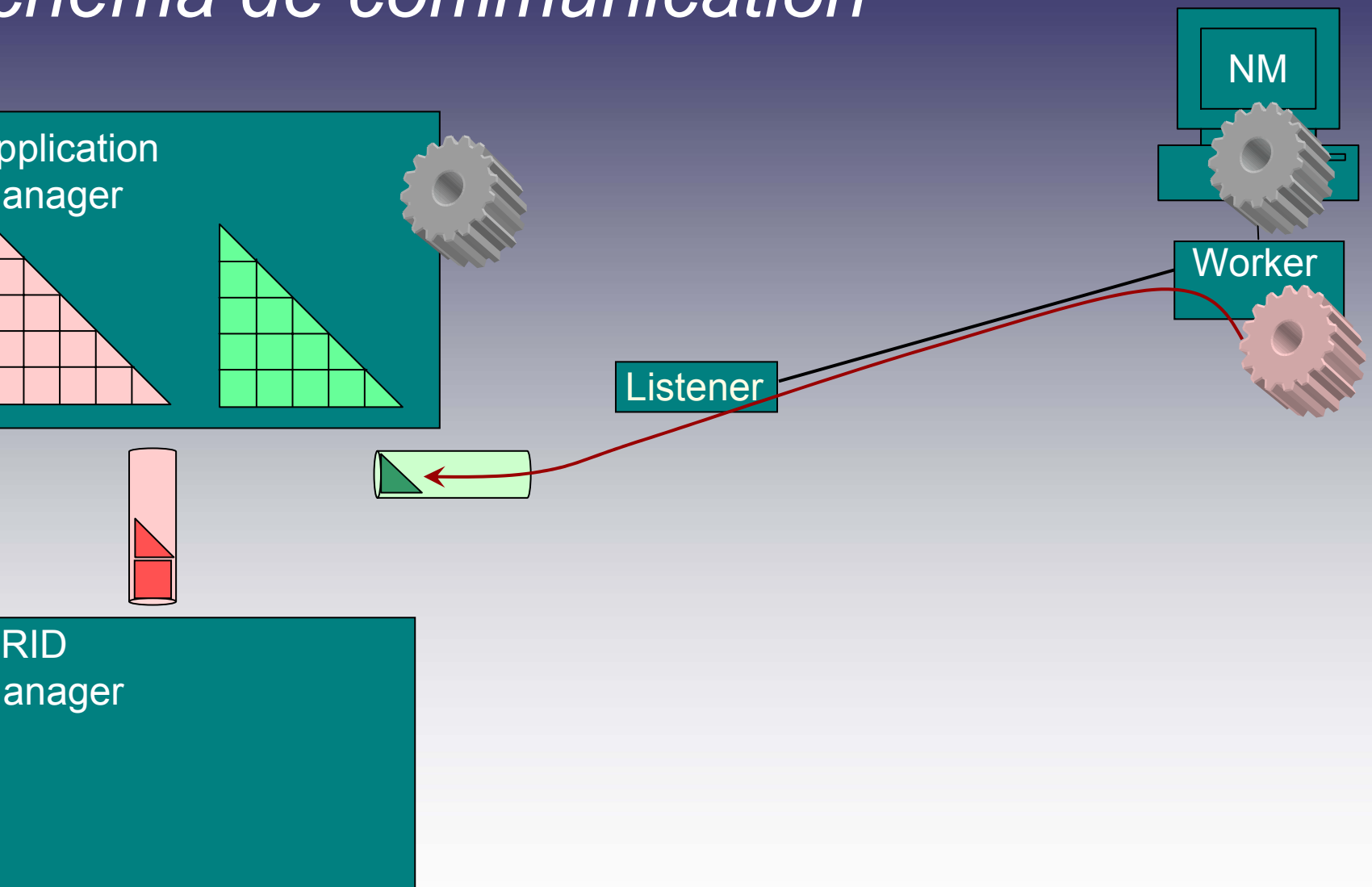
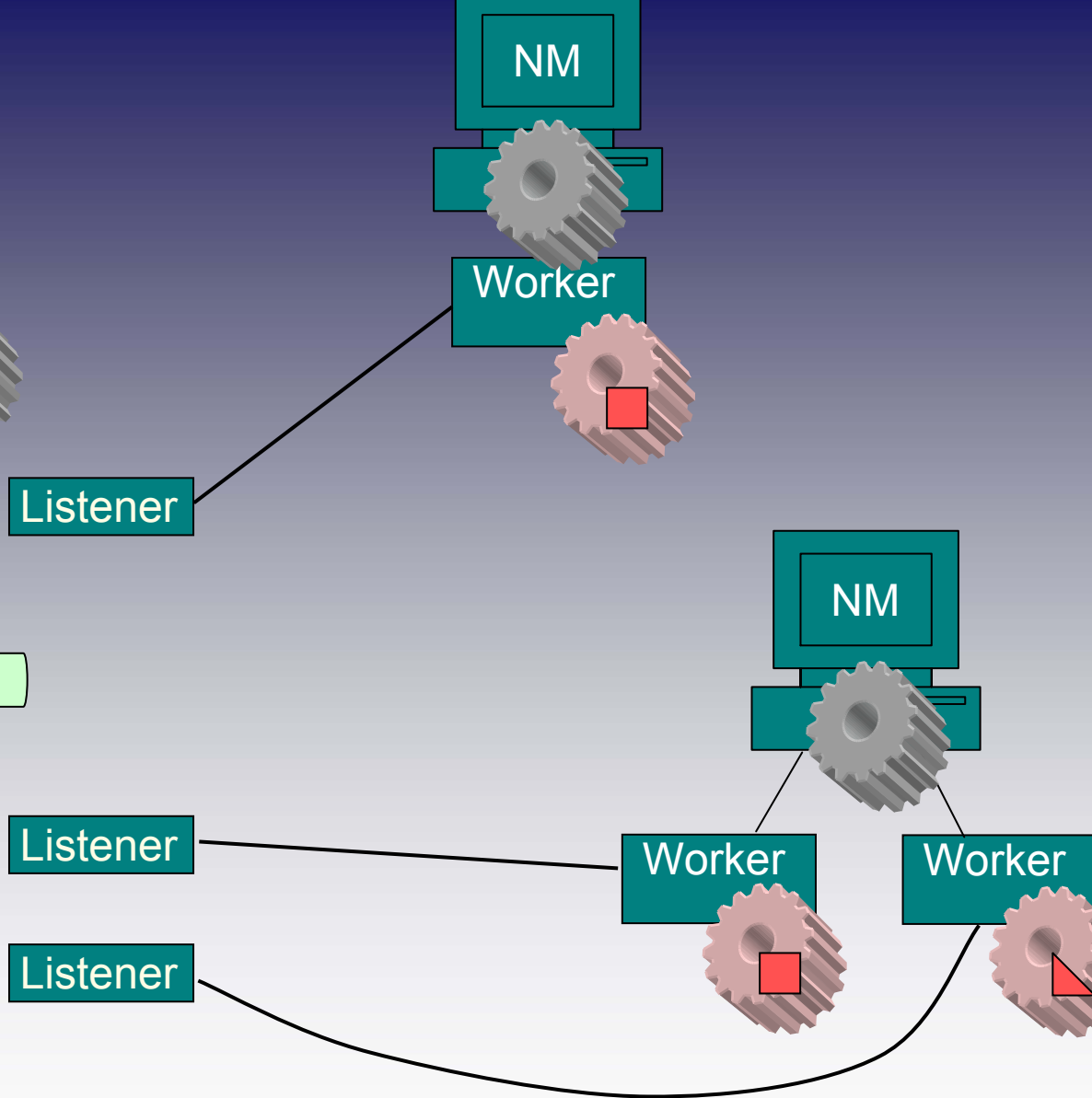
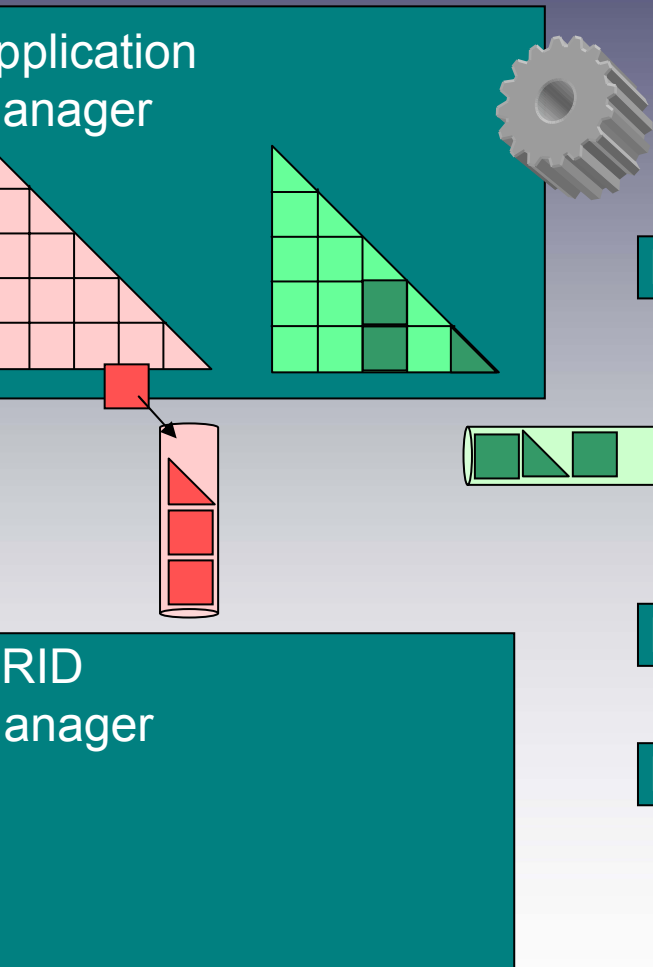


schéma de communication



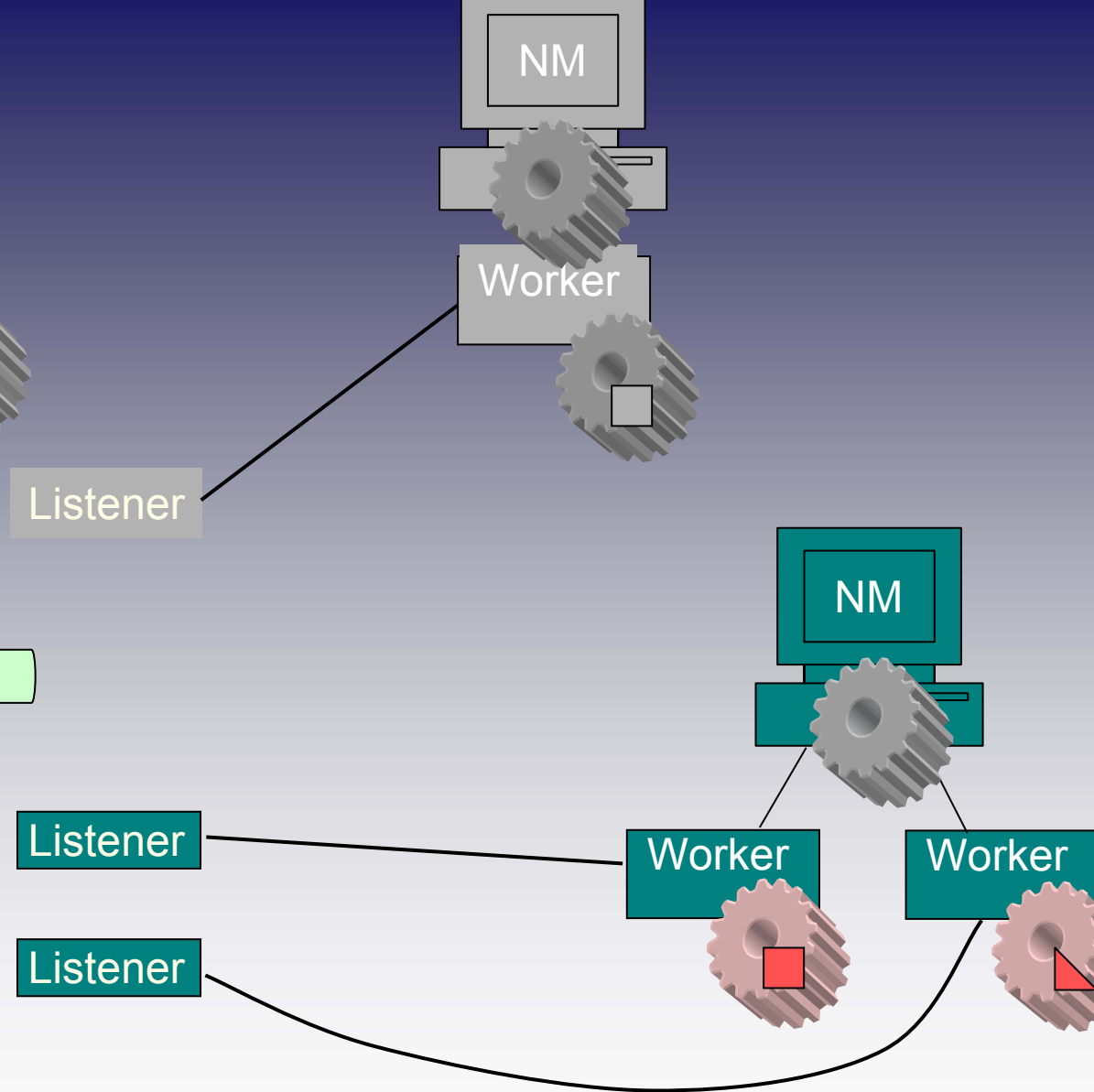
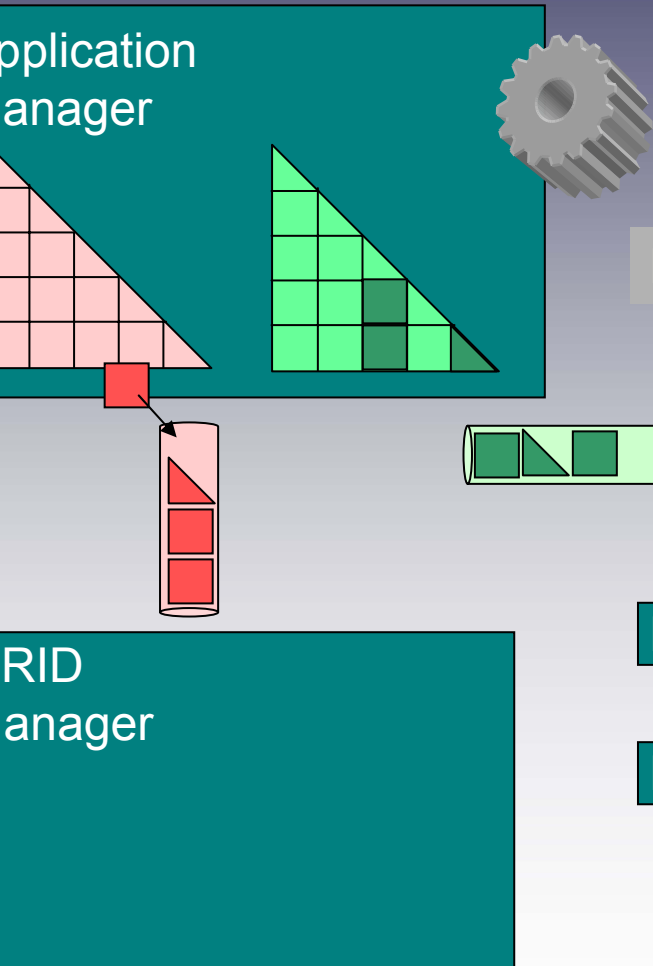
e GRID

ue globale

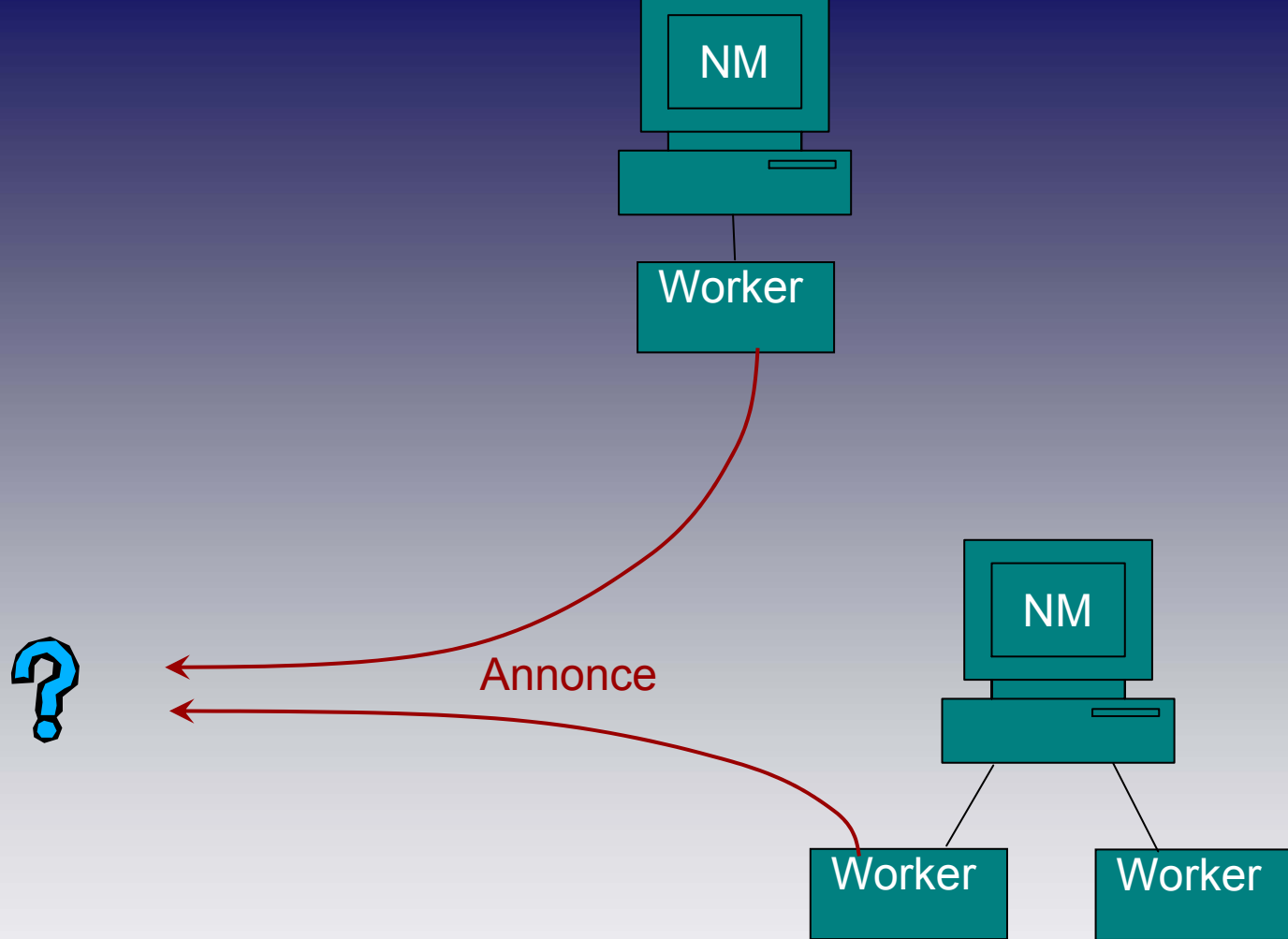


e GRID

ue globale



RID nnonces



résultats

34 Workers, 1500 signatures

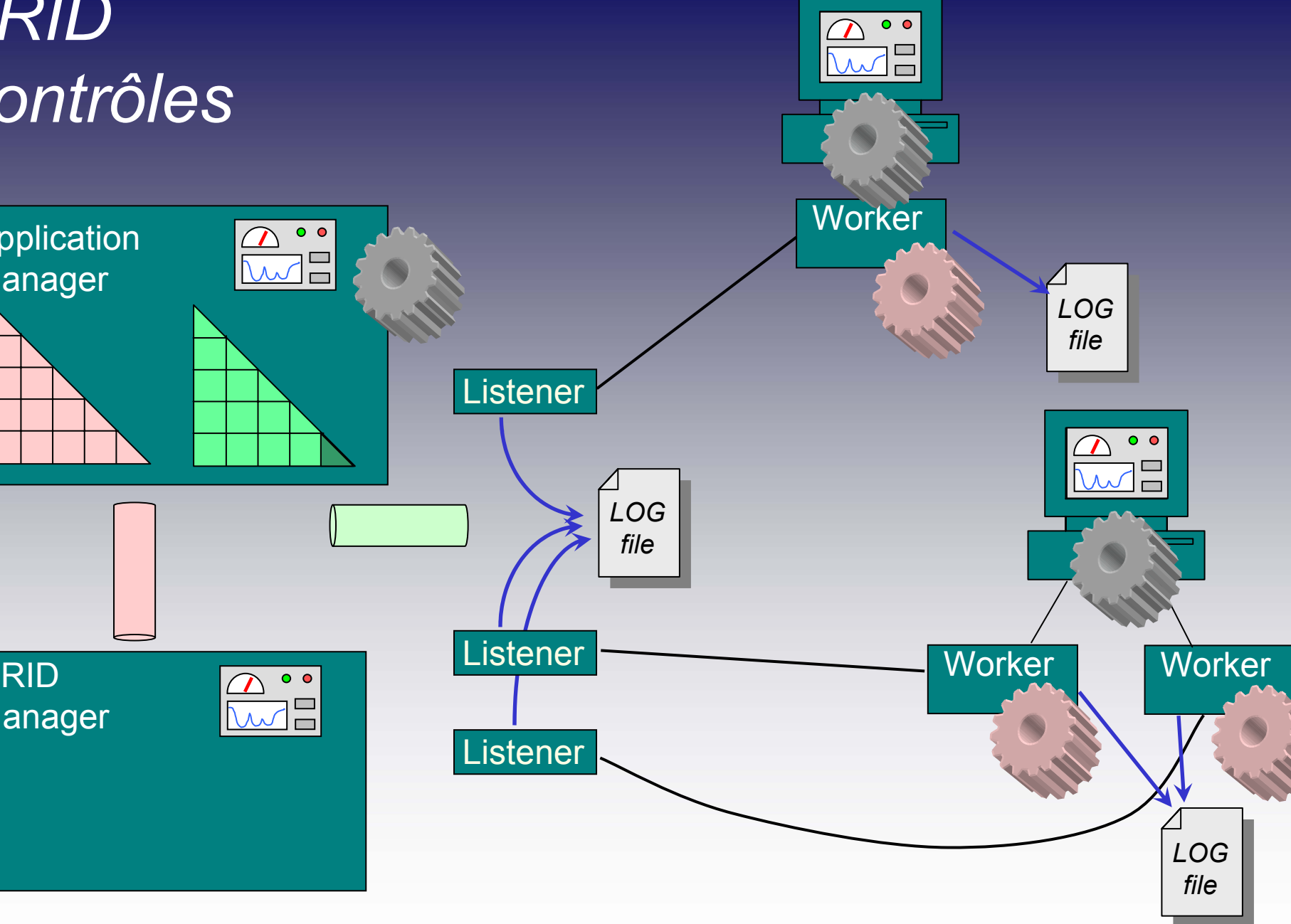
- 12 heures de calcul
- 2 semaines sur un seul processeur

Encore insuffisant

- 10 fois plus de signatures => 100 fois plus de comparaisons
- Capacité totale : 64 processeurs

⇒ Optimiser la partie « atomique » du calcul

RID Contrôles



Conclusion

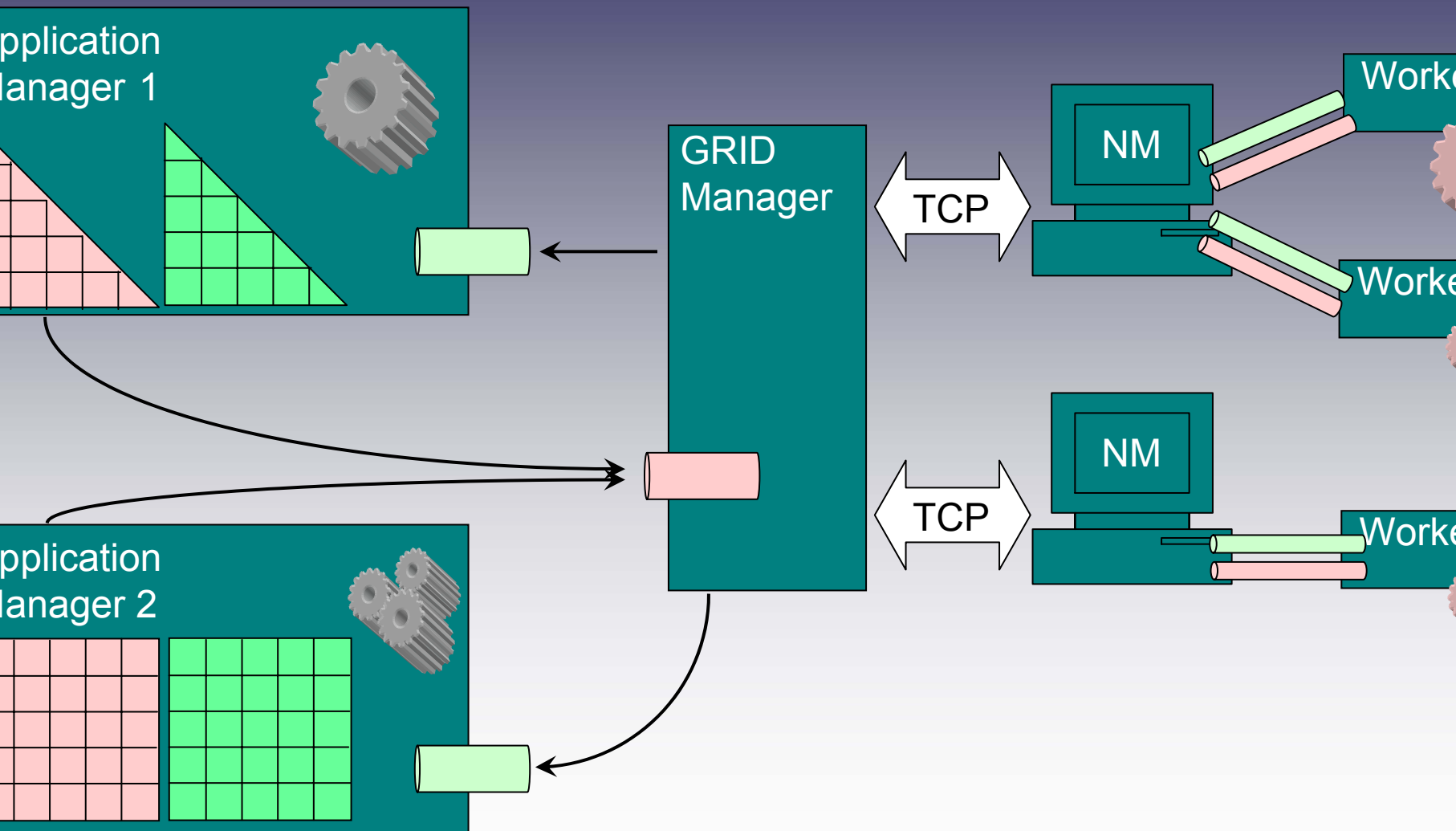
Prototype

- Faisabilité du projet
- Exploitation de la puissance de calcul

Améliorations

- Optimisation des comparaisons de signatures
- Souplesse et robustesse du GRID

FuturVERSION



A quand un *GRID Developers Toolkit* pour
LabVIEW ?