

Segmentation des usages des clients ADSL à partir de mesures de trafic

Françoise Fessant, Amine Boussarsar, Fabrice CLEROT
FTRD Lannion, TECH/SUSI/TSI

Le présent document contient des informations qui sont la propriété de France Télécom. L'acceptation de ce document par son destinataire implique, de la part de ce dernier, la reconnaissance du caractère confidentiel de son contenu et l'engagement de n'en faire aucune reproduction, aucune transmission à des tiers, aucune divulgation et aucune utilisation commerciale sans l'accord préalable écrit de France Télécom R&D

Plan :



- Objectifs
- Rappel sur les données observées
- Analyse simultanée des usages et des profils journaliers de trafic
 - Méthodologie de l'analyse, synopsis
 - Mise en œuvre sur les données OTARIE
 - Utilisation pratique des segments de clients
- Conclusion

Plan :



- Objectifs
- Rappel sur les données observées
- Analyse simultanée des usages et des profils journaliers de trafic
 - Méthodologie de l'analyse, synopsis
 - Mise en œuvre sur les données OTARIE
 - Utilisation pratique des segments de clients
- Conclusion

Objectifs de l'étude



- Analyse des usages des clients ADSL
 - Segmentation des clients en fonction de leurs usages
 - Analyse à partir des données de trafic observées
 - Complémentaire par rapport aux études de panel
 - Moins détaillé dans l'analyse des usages
 - Plus fiable statistiquement

- Meilleure compréhension des usages
- Meilleure anticipation des besoins



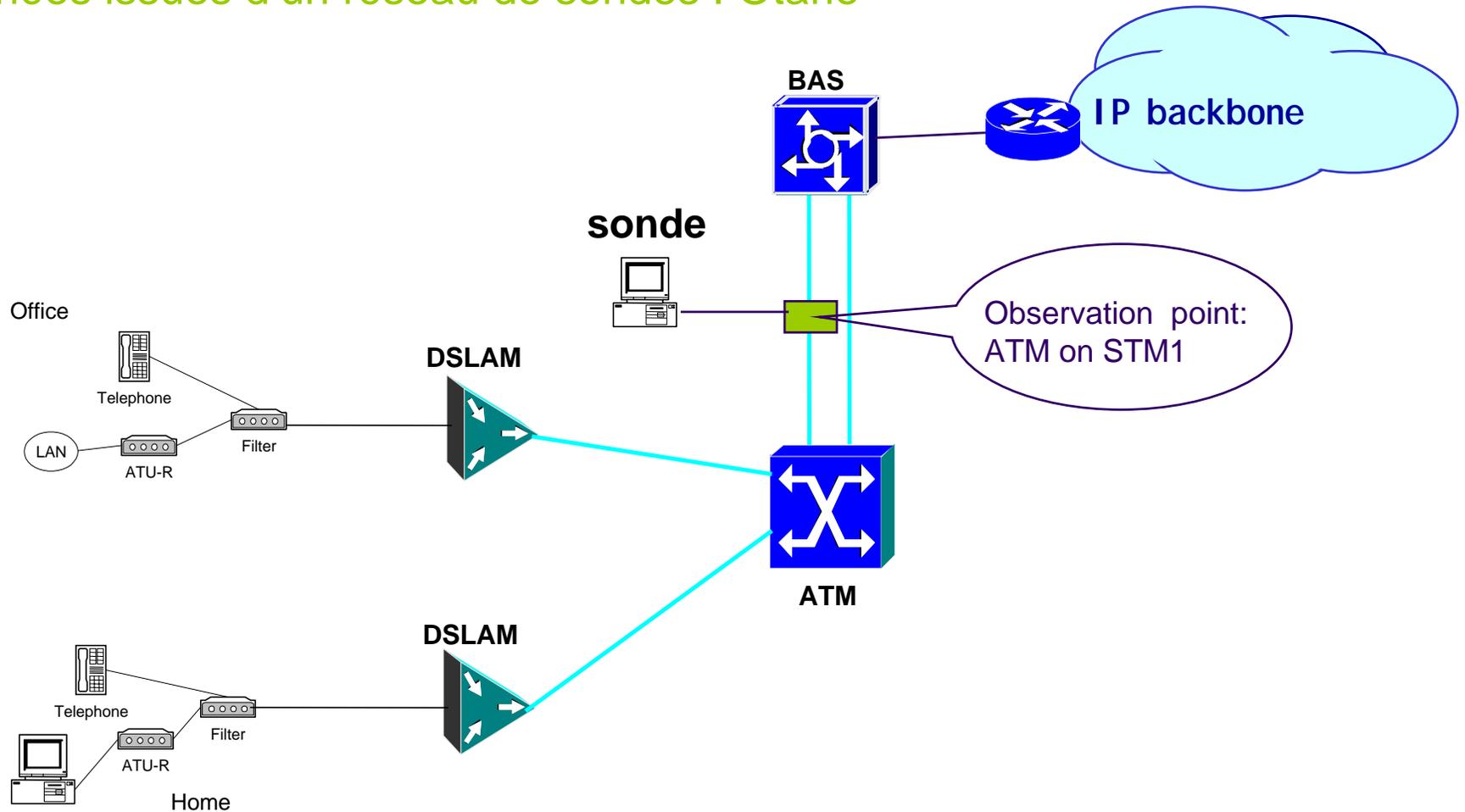
Plan :

- Objectifs
- Rappel sur les données observées
- Analyse simultanée des usages et des profils horaires de trafic
 - Méthodologie de l'analyse, synopsis
 - Mise en œuvre sur les données OTARIE
 - Utilisation pratique des segments de clients
- Conclusion



Rappel sur les données observées

→ Données issues d'un réseau de sondes : Otarie





Rappel sur les données observées

- Données issues d'un réseau de sondes : Otarie
- Trafics regroupés sur 13 catégories de ports
 - Unknown, web, P2P, FTP, News, Mail, DB, others, control, games, streaming, chat, VOIP
- Campagne de mesure sur le long terme :
 - 5 sites géographiques
 - début des mesures : janvier 2003
- Format des données :
 - chaque jour, pour chaque client, les volumes échangés dans les sens de trafic montant et descendant sur les 13 catégories par plage de 6'

Plan :



- Objectifs
- Rappel sur les données observées
- Analyse simultanée des usages et des profils horaires de trafic
 - Méthodologie de l'analyse, synopsis
 - Mise en œuvre sur les données OTARIE
 - Utilisation pratique des segments de clients
- Conclusion

Plan :



- Objectifs
- Rappel sur les données observées
- Analyse simultanée des usages et des profils horaires de trafic
 - Méthodologie de l'analyse, synopsis
 - Mise en œuvre sur les données OTARIE
 - Utilisation pratique des segments de clients
- Conclusion



Analyse simultanée des usages et des profils horaires de trafic

- Données de départ : pour chaque client, pour chaque jour, sur un mois, le profil horaire du volume échangé sur chaque application

- Organisation des données en 3 niveaux de hiérarchie
 - Chaque client est décrit par ses journées
 - Chaque journée est décrite par ses applications
 - Chaque application est décrite par son volume horaire de trafic

fichier log

Client1	jour1	email	volumejoureemail11
Client1	jour1	P2P	volumejourP2P11
Client1	jour2	email	volumejoureemail12
Client1	jour2	web	volumejourweb12
Client2	jour1	email	volumejoureemail21
Client2	jour3	email	volumejoureemail23
Client2	jour3	chat	volumejourchat23
Client2	jour3	games	volumejourgames23
Client2	jour5	P2P	volumejourP2P25
Client3	jour2	P2P	volumejourP2P32
Client3	jour2	web	volumejourweb32
...			



Analyse simultanée des usages et des profils horaires de trafic

→ Analyse exploratoire en plusieurs étapes :

- Segmentation de tous les profils journaliers d'activité *tous clients confondus*
 - Segmentation de tous les profils journaliers par application
Découverte de jours/applications « typiques »

 - Rendre à chaque jour, pour chaque application, « ses » profils d'activité jour/application et représenter chaque jour sur la base des segments jour/application typiques

 - Segmentation des jours sur la base de cette représentation
Découverte des jours d'activité « typiques »

- Rendre à chaque client « ses » profils journaliers d'activité et représenter chaque client par sa fréquence sur chaque segment de profils journaliers

- Segmentation des clients sur la base de cette représentation
Découverte des clients « typiques »



Analyse simultanée des usages et des profils horaires de trafic

→ Analyse exploratoire en plusieurs étapes :

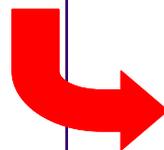
- Segmentation de tous les profils journaliers d'activité *tous clients confondus*
 - Segmentation de tous les profils journaliers par application
Découverte de jours/applications « typiques »
 - Rendre à chaque jour pour chaque application « ses » profils d'activité jour/application et représenter chaque jour sur la base des segments jour/application typiques
 - Segmentation des jours sur la base de cette représentation
Découverte des jours d'activité « typiques »
- Rendre à chaque client « ses » profils journaliers d'activité et représenter chaque client par sa fréquence sur chaque segment de profils journaliers
- Segmentation des clients sur la base de cette représentation
Découverte des clients « typiques »

Découverte des jours « typiques » par application



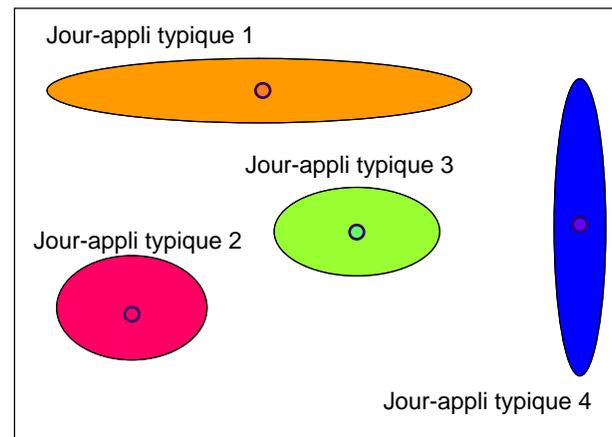
fichier log

Client1	jour1	email	volumejouremail11
Client1	jour1	P2P	volumejourP2P11
Client1	jour2	email	volumejouremail12
Client1	jour2	web	volumejourweb12
Client2	jour1	email	volumejouremail21
Client2	jour3	email	volumejouremail23
Client2	jour3	chat	volumejourchat23
Client2	jour3	games	volumejourgames23
Client2	jour5	P2P	volumejourP2P25
Client3	jour2	P2P	volumejourP2P32
Client3	jour2	web	volumejourweb32
...			

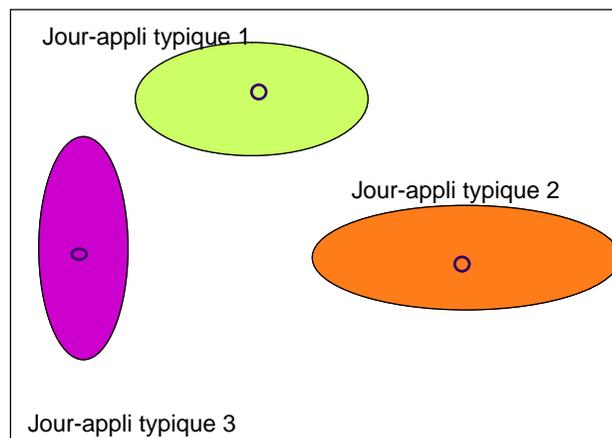


Clustering des jours-applications
dans l'espace des profils horaires de trafic

Espace des volumes horaires, email



Espace des volumes horaires, P2P



Segments de jours typiques, pour chaque application,
décrits sur la base des profils horaires de trafic
(pour l'application)



Clustering des jours-applications
dans l'espace des profils horaires de trafic



Analyse simultanée des usages et des profils horaires de trafic

→ Analyse exploratoire en plusieurs étapes :

- Segmentation de tous les profils journaliers d'activité *tous clients confondus*
 - Segmentation de tous les profils journaliers par application
Découverte de jours/applications « typiques »

 - Rendre à chaque jour, pour chaque application, « ses » profils d'activité jour/application et représenter chaque jour sur la base des segments jour/application typiques

 - Segmentation des jours sur la base de cette représentation
Découverte des jours d'activité « typiques »

- Rendre à chaque client « ses » profils journaliers d'activité et représenter chaque client par sa fréquence sur chaque segment de profils journaliers

- Segmentation des clients sur la base de cette représentation
Découverte des clients « typiques »

Projection des jours sur les jours-applications typiques

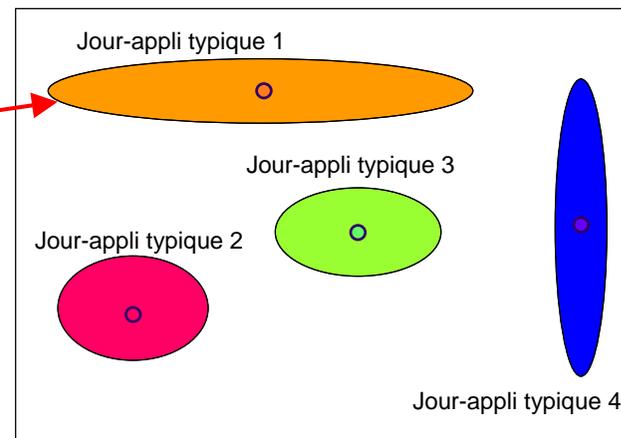


Profil de la journée pour l'application email

fichier log

Client1	jour1	email	volumejouemail11
Client1	jour1	P2P	volumejourP2P11
Client1	jour2	email	volumejouemail12
Client1	jour2	web	volumejourweb12
Client2	jour1	email	volumejouemail21
Client2	jour3	email	volumejouemail23
Client2	jour3	chat	volumejourchat23
Client2	jour3	games	volumejougames23
Client2	jour5	P2P	volumejourP2P25
Client3	jour2	P2P	volumejourP2P32
Client3	jour2	web	volumejourweb32
...			

1 0 0 0

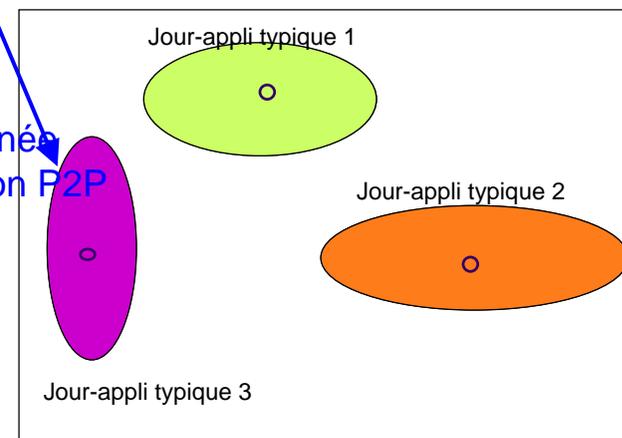


0 0 1

Profil de la journée pour l'application P2P

Profils de journées

Client1	jour1	1 0 0 0	0 0 1	0 0 0 0 0	...
Client1	jour2				
...					



- Chaque journée est affectée à un cluster pour chaque application
- Une journée est caractérisée par la concaténation des profils jours-applications typiques (corrélation des applications au niveau de la journée)



Analyse simultanée des usages et des profils horaires de trafic

→ Analyse exploratoire en plusieurs étapes :

- Segmentation de tous les profils journaliers d'activité *tous clients confondus*
 - Segmentation de tous les profils journaliers par application
Découverte de jours/applications « typiques »

 - Rendre à chaque jour, pour chaque application, « ses » profils d'activité jour/application et représenter chaque jour sur la base des segments jour/application typiques

 - Segmentation des jours sur la base de cette représentation
Découverte des jours d'activité « typiques »

- Rendre à chaque client « ses » profils journaliers d'activité et représenter chaque client par sa fréquence sur chaque segment de profils journaliers

- Segmentation des clients sur la base de cette représentation
Découverte des clients « typiques »

Découverte des jours d'activité « typiques »

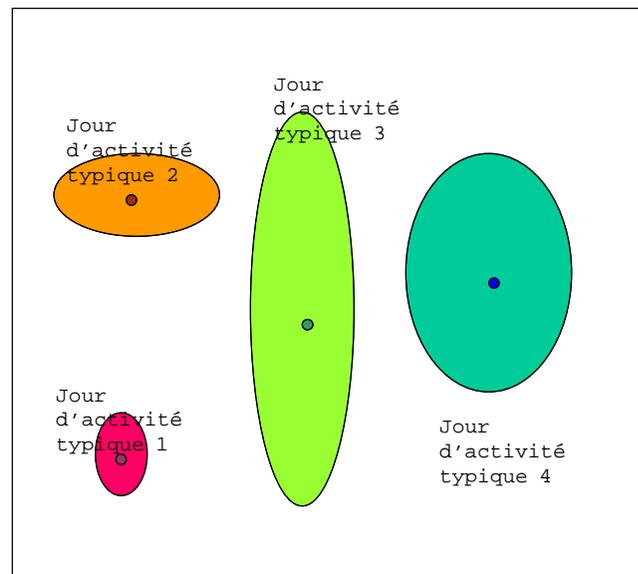


fichier log

Client1	jour1	Profiljour11
Client1	jour2	Profiljour12
Client2	jour1	ProfilJour21
Client2	jour3	ProfilJour23
Client2	jour5	ProfilJour25
Client3	jour2	ProfilJour32
Client4	jour4	ProfilJour44
Client4	jour5	ProfilJour45
Client5	jour6	ProfilJour56
...		



Espace des profils jours/applications



- Segments de jours d'activité typiques décrits sur la base des segments jours-applications



Analyse simultanée des usages et des profils horaires de trafic

→ Analyse exploratoire en plusieurs étapes :

- Segmentation de tous les profils journaliers d'activité *tous clients confondus*
 - Segmentation de tous les profils journaliers par application
Découverte de jours/applications « typiques »

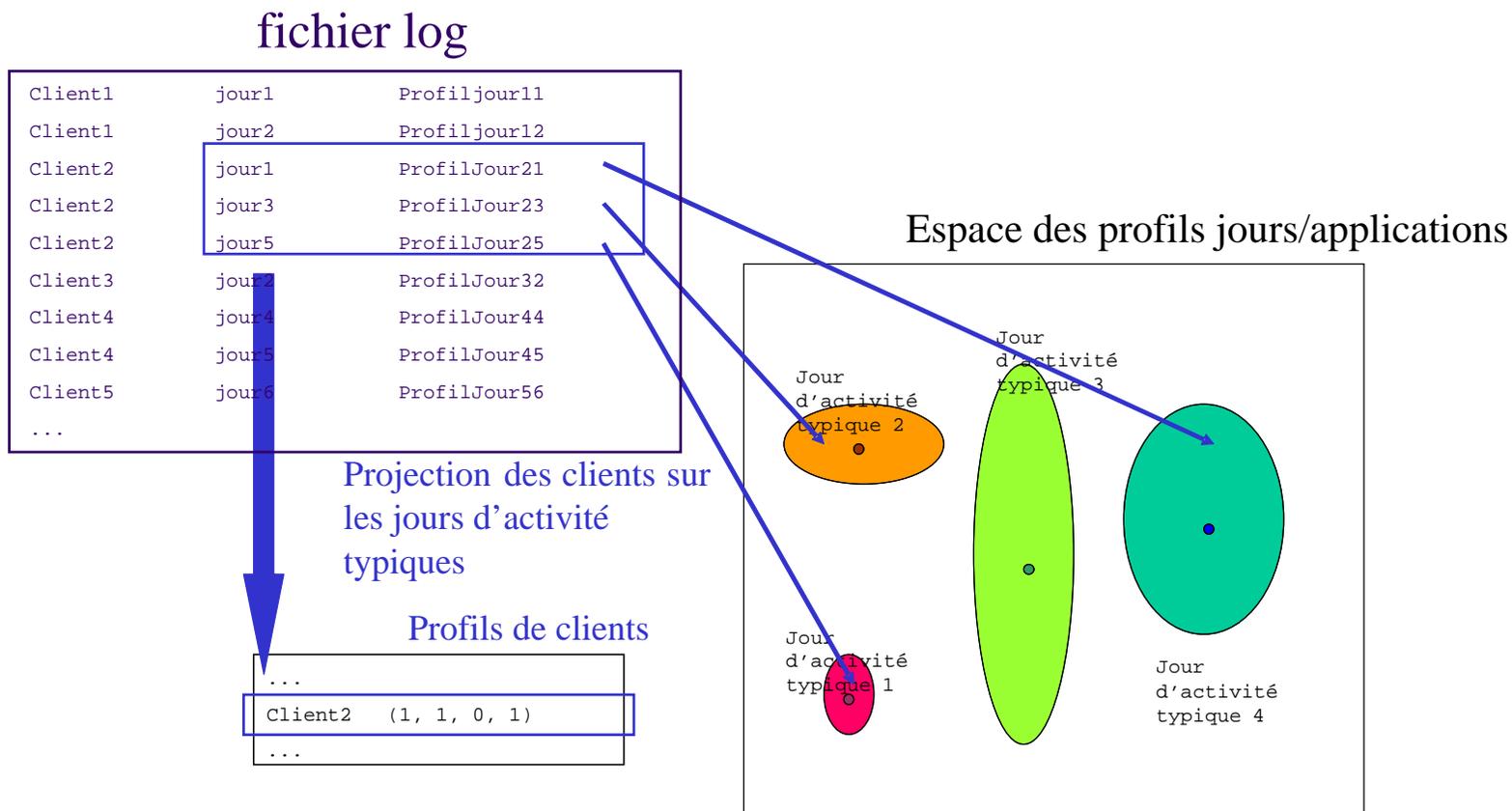
 - Rendre à chaque jour, pour chaque application, « ses » profils d'activité jour/application et représenter chaque jour sur la base des segments jour/application typiques

 - Segmentation des jours sur la base de cette représentation
Découverte des jours d'activité « typiques »

- Rendre à chaque client « ses » profils journaliers d'activité et représenter chaque client par sa fréquence sur chaque segment de profils journaliers

- Segmentation des clients sur la base de cette représentation
Découverte des clients « typiques »

Projection des clients sur les jours d'activité typiques



- Chaque journée est affectée à un cluster
- Un client est caractérisé par un profil de journées d'activités typiques



Analyse simultanée des usages et des profils horaires de trafic

→ Analyse exploratoire en plusieurs étapes :

- Segmentation de tous les profils journaliers d'activité *tous clients confondus*
 - Segmentation de tous les profils journaliers par application
Découverte de jours/applications « typiques »

 - Rendre à chaque jour, pour chaque application, « ses » profils d'activité jour/application et représenter chaque jour sur la base des segments jour/application typiques

 - Segmentation des jours sur la base de cette représentation
Découverte des jours d'activité « typiques »

- Rendre à chaque client « ses » profils journaliers d'activité et représenter chaque client par sa fréquence sur chaque segment de profils journaliers

- Segmentation des clients sur la base de cette représentation
Découverte des clients « typiques »

Découverte des clients « typiques »

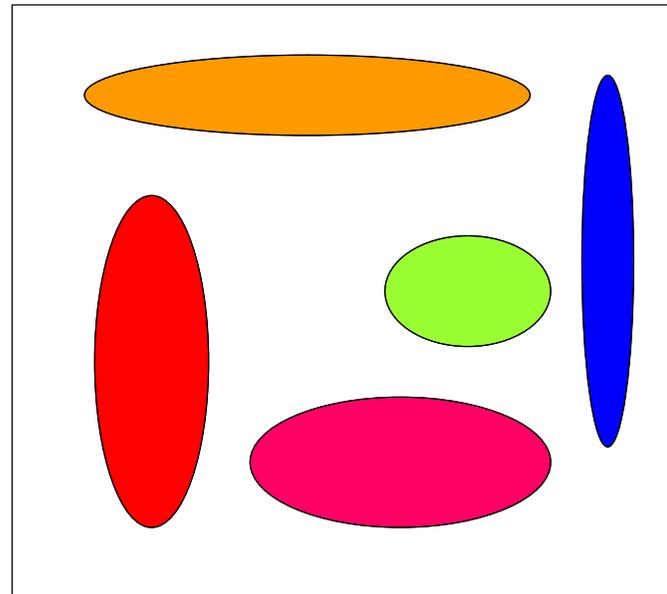


Client1	ProfilClient1
Client2	ProfilClient2
Client3	ProfilClient3
...	



Clustering des clients dans
l'espace des jours d'activités
typiques

Espace des jours d'activités typiques



- Segments de clients typiques décrits sur la base des jours d'activité typiques



Plan :

- Objectifs
- Rappel sur les données observées
- Analyse simultanée des usages et des profils horaires
 - Méthodologie de l'analyse, synopsis
 - Mise en œuvre sur les données OTARIE
 - Utilisation pratique des segments de clients
- Conclusion



→ Site de Fontenay, Février 2004

- 1991 clients « actifs » au moins une fois sur le mois
- 3 applications qui génèrent les plus forts volumes : Unknown, Web, P2P

→ Analyse exploratoire par cartes de Kohonen

- Segmentation et visualisation de données multidimensionnelles
 - « Projection » dans un espace bidimensionnel
 - effectue une segmentation tout en préservant les relations de voisinage entre objets

→ Détail technique :

- Etant donné les très grands écarts entre profils jour-application au niveau des volumes échangés, on classe les logarithmes des profils jours-applications

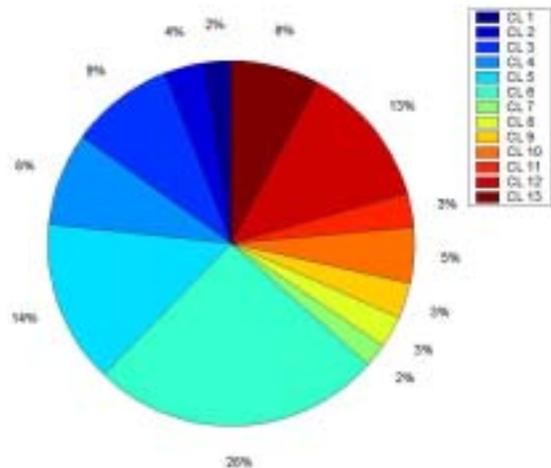
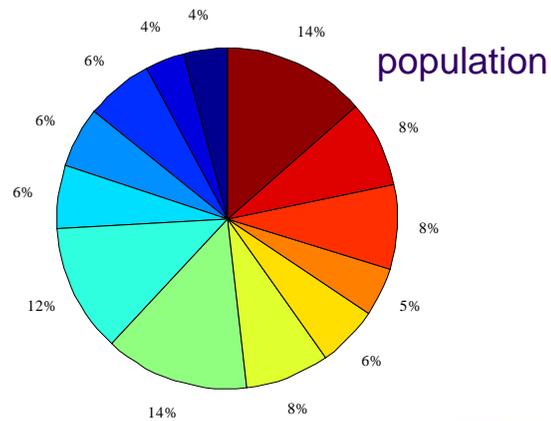
Segmentation des applications : exemple



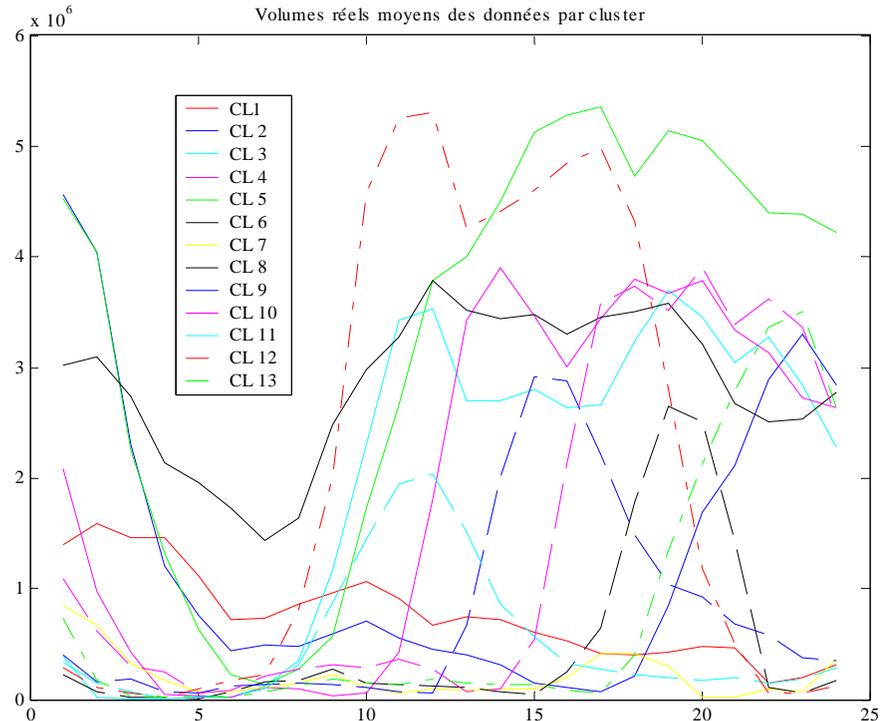
➔ Application : Web down

➔ 13 profils journaliers typiques pour l'application

- Journées inactives (14% des journées)
- Journées avec une activité piquée sur une plage horaire précise (39% des journées)
- Journées avec de longues plages d'activité (47% des journées)



volume

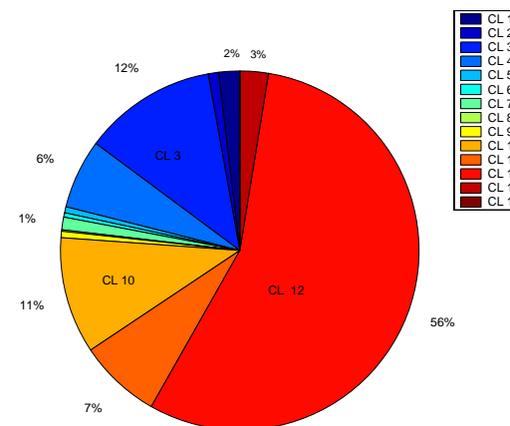
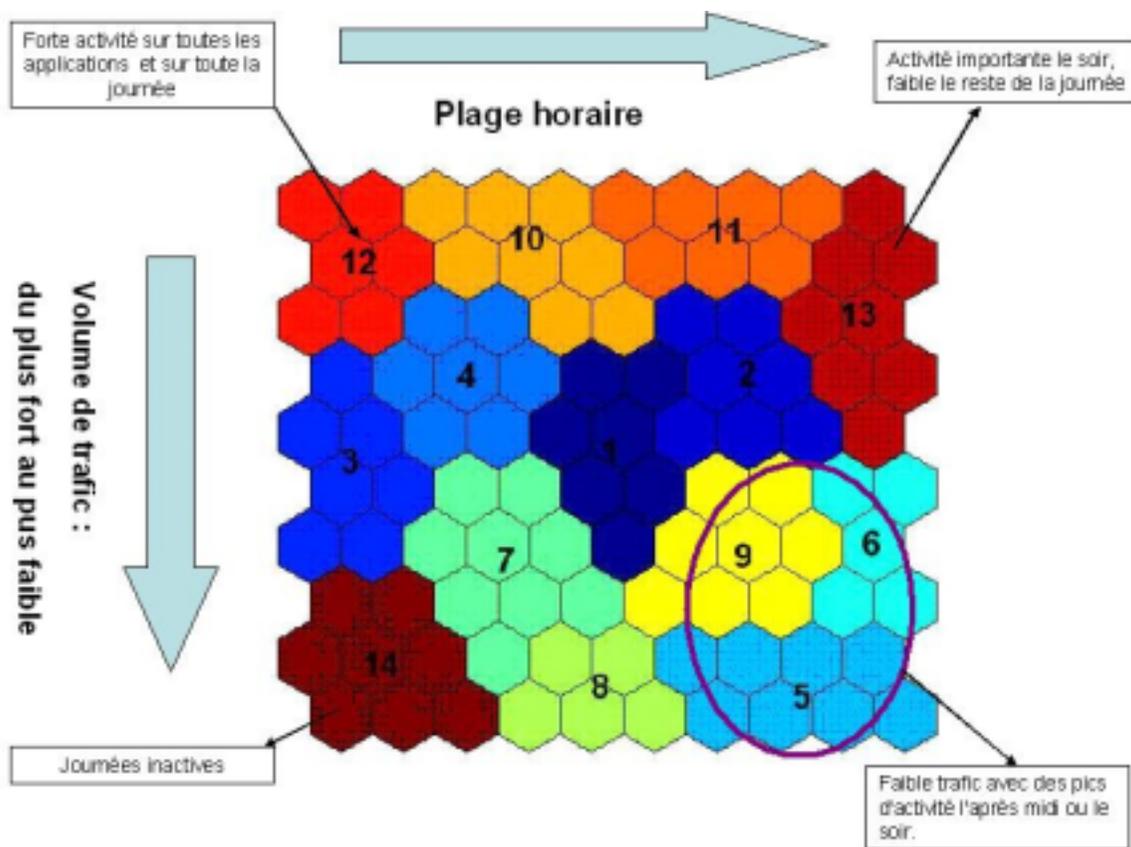


Segmentation des journées

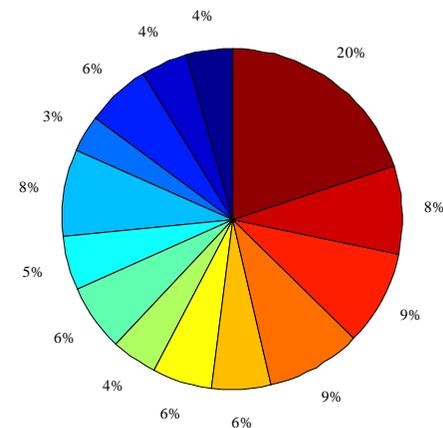


→ 14 profils journaliers typiques

- Utilisation différente des 3 applications en terme d'intensité du trafic dans un cluster, mais trafic sur les mêmes plages horaires
- Symétrie des usages up et down



volume



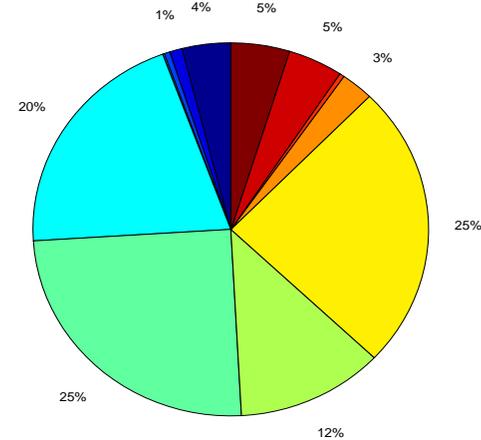
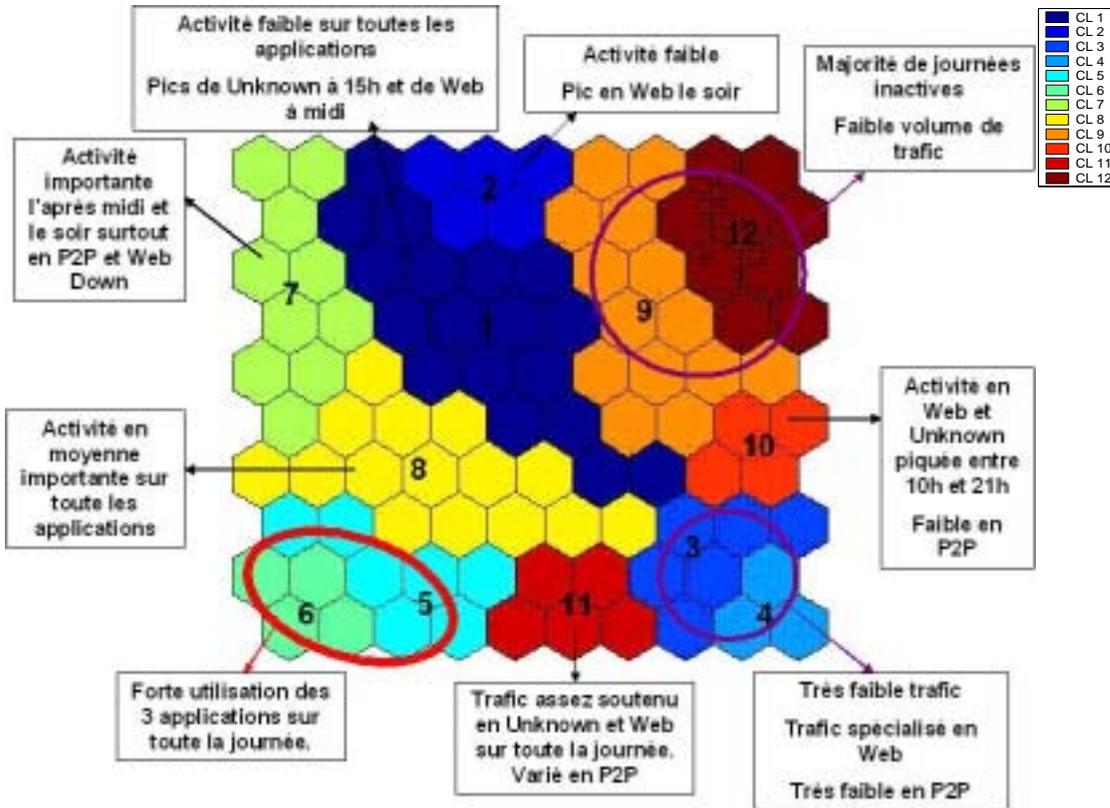
population

Segmentation des clients

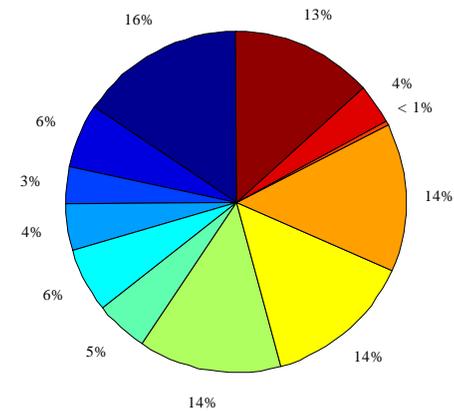


→ 12 clients typiques

➤ caractérisés en fonction des usages des applications et des plages horaires de trafic



volume



population



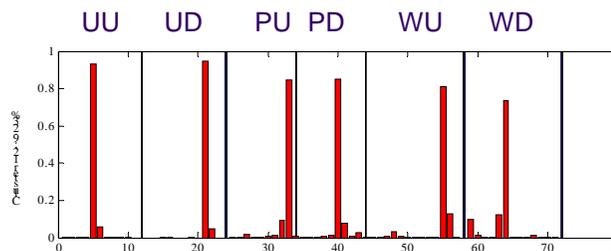
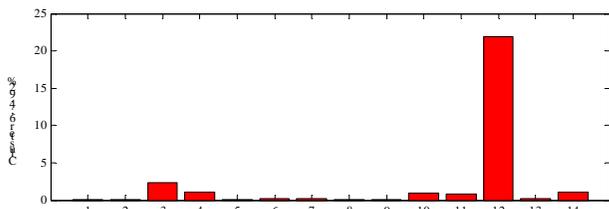
Plan :

- Objectifs
- Rappel sur les données observées
- Analyse simultanée des usages et des profils horaires
 - Méthodologie de l'analyse, principe
 - Mise en œuvre sur les données OTARIE
 - Utilisation pratique des segments de clients
- Conclusion

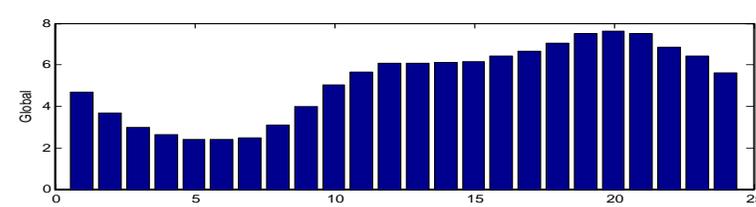
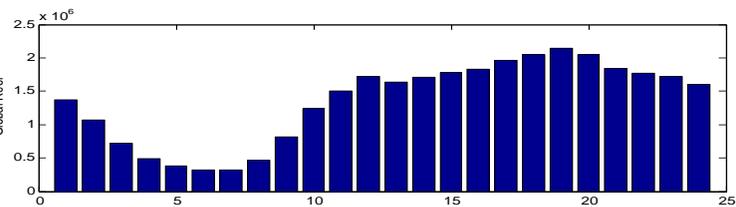
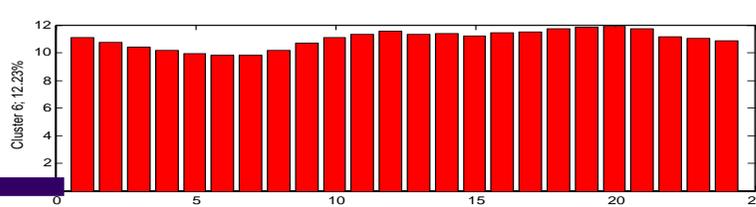
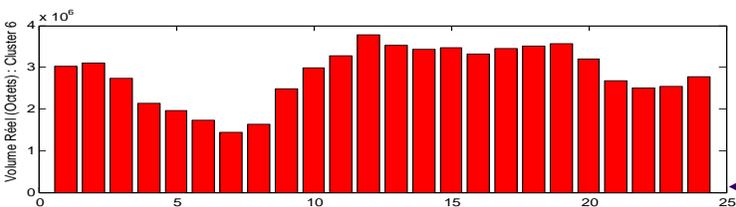
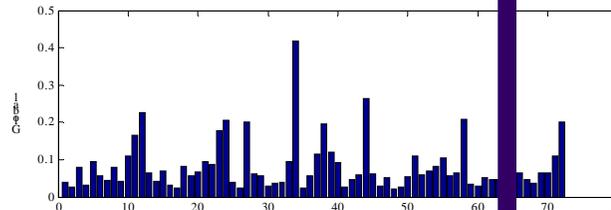
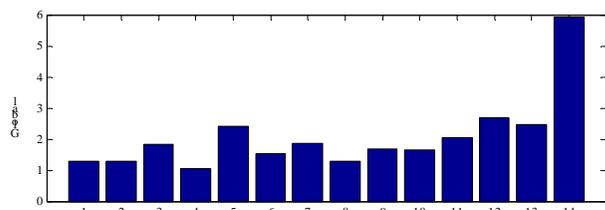
Utilisation pratique de la segmentation



80 % de journées très actives (5% des clients)
 activité en continu sur la journée pour toutes les applications UU



Journée type 12, un type dominant par application
 Web down, journée type 6 à 85%



profil moyen du cluster, volume réel

fort trafic en continu sur la journée
 (profil moyen du cluster, moyenne des logs)

→ Segmentation fine des clients ADSL

- Segmentation simultanée des profils journaliers de toutes les catégories de ports
- Objectif :
 - avoir une vue simultanée des usages et des profils horaires
 - « qui fait quoi et quand ? »

→ Méthodologie d'analyse exploratoire qui exploite la structure hiérarchique des données

- Basée sur les cartes de Kohonen
- Segments de clients basés sur une description agrégée de leur activité journalière
- Segments simples à analyser et à interpréter
- Laisse la liberté de « détailler » les segments à tout moment selon les besoins

Projection d'informations externes



→ lien entre les usages et le type de contrat netissimo

