

**République Algérienne Démocratique et Populaire**  
**Ministère de l'Enseignement Supérieure de la Recherche Scientifique**  
**Université Dr MOULAY Tahar de Saïda**



**Faculté des Technologies**  
**Département d'Electronique**



**MEMOIRE DE FIN D'ETUDES**  
**Master en Réseaux & Télécommunication**  
**Thème :**

---

**Contribution à l'évolution de système DSCDMA en  
utilise technique de DATA MINING**

---

Présenté par :

**AMAM Mustapha**

**BOUCETTA Mohamed**

<b>N°</b>	<b>Jury</b>	<b>Nom Et Prénom</b>	<b>Grade</b>	<b>Université</b>
<b>01</b>	Président	Boudkhile Abdelhakime	Dr	Université Dr Tahar Moulay de Saida
<b>02</b>	Rapporteur	Bouasria Fatima	Pr	Université Dr Tahar Moulay de Saida
<b>03</b>	Examineur	Meddah Ishak	Dr	Université Dr Tahar Moulay de Saida

**Année universitaire : 2019/2020**



# DEDICACE

JE DEDIE CE MEMOIRE

A la mémoire de mes parents  
A mes frères, sœurs et leurs petites familles A  
mes amies que j'adore

A tous ceux qui m'ont aidé de près ou de loin A  
mon encadreur Mme Bouasria fatima

**AMAM Mustapha**

# DEDICACE

JE DEDIE CE MEMOIRE

A la mémoire de mes parents  
A mes Frères, Sœurs et leurs petites familles A  
mes amies que j'adore

A tous ceux qui m'ont aidé de près ou de loin A  
mon encadreur Mme Bouasria fatima

**BOUCETTA Mohamed**

# Remerciement

**Tout d'abord, on tient à remercier le bon Dieu le tout Puissant de nous avoir donné la force et le courage de mener à bien ce modeste travail, également on remercie infiniment nos parents, qui nous ont encouragé à arriver à ce stade de notre formation.**

**On tient a remercier tous ceux et celles qui ont contribué finaliser ce modeste travail.**

**Nos remerciements vont à docteur Bouasria fatima notre encadreur pour nous avoir guider pour la réalisation de ce projet.**

**AMAM Mustapha & BOUCETTA Mohamed**



## Résumé

L'étalement de spectre est une technique très robuste pour combattre les interférences causées par les brouilleurs ou les interférences d'accès multiples. Cette technique est à la base des systèmes radio-mobiles de 3<sup>ème</sup> génération.

Ce projet de fin d'étude est consacré à la mise en évidence des performances des systèmes à étalements de spectre dans une chaîne de communication multiutilisateurs avec un canal multi-trajets.

**Mots clés :** Canal radio-mobile, étalement de spectre, communications numériques, accès multiples, brouilleurs.

---

## المخلص

النطاق الترددي هي تقنية قوية جدا لمكافحة التداخلات التي تسببها أجهزة التشويش أو تداخلات الوصل المتعدد بسبب تبادل الاتصالات بين المستخدمين في نفس الوقت. هذه التقنية هي التي كانت وراء تطور أنظمة الاتصالات للجيل الثالث. مشروع نهاية الدراسة هذا يتمحور حول دراسة أداء ومدى كفاءة أنظمة النطاق الترددي العريض في سلسلة اتصال متعدد المستخدمين عبر قناة اتصال متعددة المسالك

**كلمات مفتاحية:** قناة الاتصال، نطاق الترددي، اتصال الرقمي، الوصول المتعدد

---

## Abstract

The spread spectrum is a technique very efficient to combat the interference caused by the Jammers or the interference of multiple access. This technique is at the appearance of the mobile radio systems of 3rd generation.

This project is devoted to the highlighting of the performance of the spread spectrum systems multiuser communication chain via multiple paths channel.

**Keywords:** radio mobile channel, spread spectrum, digital communication, multiple access,

## Sommaire :

### Chapitre 1 :

#### Généralités sur les Réseaux cellulaires

I.1	Introduction	1	
I.2	Concept cellulaire	1	
I.2.1	Type des cellules	1	
1.3	Évolution des réseaux cellulaires	2	
I.3.1	La première génération (1G)	2	
I.3.2	La deuxième génération 2G (GSM)	2	2
I.3.3	Evolution vers l'EDGE	3	
I.3.4	La troisième génération (3G)	3	
I.4	Présentation du réseau GSM	3	
I.4.1	Architecture du réseau GSM	3	
1.4.1.1	Station mobile (MS)	4	
1.4.1.2	Le sous-système radio (BSS)	4	
1.4.1.3	Le sous-système réseau (NSS)	5	
1.4.1.4	Le Sous System d'Exploitation et de Maintenance (OSS)	6	
1.5	Présentation d'un réseau UMTS	7	
I.5.1	L'architecture du réseau UMTS	7	
I.5.1.1	L'équipement usager UE	8	
I.5.1.2	Le réseau d'accès UTRAN	8	
I.5.2	Organisation des fréquences dans un réseau UMTS	10	
I.5.3	Organisation temporelle	10	
I.5.4	La technique d'accès dans un réseau UMTS	11	
1.5.4.1	Accès multiple par répartition dans le temps (TDMA)	11	
1.5.4.2	Accès multiple par répartition de fréquences (FDMA)	13	
1.5.4.3	Accès multiple par répartition de code (CDMA)	14	
1.6	Conclusion		

## **Chapitre 2 :**

### **Principe de l'étalement de spectre par séquence directe 'DS- CDMA'**

2.1	Introduction	17
2.2	L'étalement de spectre par séquence directe	17
2.3	Pourquoi étaler le spectre	18
2.3.1	Exemple d'application	19
2.4	Présentation formelle de l'étalement de spectre	21
2.4.1	Elaboration analytique du signal temporel d'étalement	22
2.4.2	Elaboration analytique du signal temporel étalé	24
2.5	Les codes d'étalement	25
2.5.1	Codes orthogonaux	25
2.5.1.1	Génération de codes Walsh	25
2.5.1.2	Canalisation en utilisant les codes Walsh	28
2.6	Conclusion	31

## **Chapitre 3**

### **Data mining et Benchmark de données**

3.1	Introduction	32
3.2	Définition du data mining	32
3.3	Processus de data mining	33
3.3.1	Définition et compréhension du problème	33
3.3.2	Collecte des données	33
3.3.3	Prétraitement	34
3.3.4	Estimation du modèle	34
3.3.5	Interprétation du modèle et établissement des conclusions	34

# Table de matière

---

3.4	Technique du Data Mining	35
3.4.1	Techniques supervisées	35
3.4.1.1	Arbre de décision	35
3.4.1.2	Règles de décision	36
3.4.1.3	Régression	36
3.4.1.4	Réseaux de neurone	36
3.4.2	Techniques non supervisées	37
3.5	LE BENCHMARKING	38
3.5.1	Introduction	38
3.5.2	Définition du Benchmarking	38
3.5.3	Le concept de Benchmarking	39
3.5.4	les outils de développements	39
3.5.5	Résultat de simulation	43
3.5.5.1	Collecte et préparation des données	43
3.5.5.2	.Les ensemble de donnés utiliser dans notre benchmark	45
3.5.5.3	.Commentaire sue le figure 3.5 et 3.6	46
3.5.6	.Interprétation des resultants	50
3.5.7	.L'Arbre de Décision	50
3.5.8	.Analyse de l'arbre	51
3.6	.Conclusion	53
	Conclusion Générale	54



## Liste de figures :

<b>Figure 1.1</b>	Types de cellules.....	2
<b>Figure 1.2</b>	Schéma en bloc des différentes entités du réseau GSM.....	4
<b>Figure 1.3</b>	Architecture globale du réseau UMT.....	8
<b>Figure 1.4</b>	Architecture matérielle de réseau UTRAN du réseau UMT.....	9
<b>Figure 1.5</b>	La répartition des fréquences d'un réseau UMTS.....	10
<b>Figure 1.6</b>	La trame temporelle dans un système UMTS.....	11
<b>Figure 1.7</b>	Schéma d'un multiplexage temporel (TDMA).....	12
<b>Figure 1.8</b>	Schéma d'un multiplexage fréquentiel (FDMA).....	13
<b>Figure 1.9</b>	Principe de multiplexage en fréquence .....	14
<b>Figure 1.10</b>	Technique d'accès multiple CDMA .....	15
<b>Figure 1.11</b>	Différent technique CDMA .....	15
<b>Figure 2.1</b>	Exemple d'étalement par séquence directe.....	20
<b>Figure 2.2</b>	Schéma de principe d'un système de transmission CDMA.....	22
<b>Figure 2.3</b>	Elaboration d'une séquence d'étalement (code).....	23
<b>Figure 2.4</b>	Elaboration d'un signal étalé.....	24
<b>Figure 3.1</b>	Processus de Data Mining.....	33
<b>Figure 3.2</b>	Chargement d'une Base de données sur Weka .....	41
<b>Figure 3.3</b>	Phase d'apprentissage et test .....	42
<b>Figure 3.4</b>	préparation les données de benchmark pour le Weka.....	44
<b>Figure 3.5</b>	Représentation de nombre d'utilisateur sur Weka .....	45
<b>Figure 3.6</b>	Représentation longueur de code sur Weka.....	45
<b>Figure 3.7</b>	Représentation Energie de bit sur Weka.....	46
<b>Figure 3.8</b>	Représentation de puissance de bruit sur Weka.....	47
<b>Figure 3.9</b>	Représentation BER (Bits Erreur Rate ) sur Weka.....	48
<b>Figure 3.10</b>	Représentation de classe (bonne / mauvaise) sur Weka .....	49
<b>Figure 3.11</b>	Arbre de décision pour notre technique d'accès DSCDMA .....	51

## Liste des tables :

<b>Tableau 1.1</b>	principales différences technologiques entre le GSM, IS-95, IS-136 et le PDC.....	4
<b>Tableau 2.1</b>	Caractéristiques de quelques standards de télécommunication.....	21
<b>Tableau 2.2</b>	la suite de séquence de chiffrement P.....	23
<b>Tableau 3.1</b>	caractéristiques des tableaux de données DSSSS.....	43

## Liste des abréviations :

<b>GSM</b>	Global System for Mobile Communications
<b>UMTS</b>	Universal Mobile Telecommunications System (UMTS)
<b>1 G</b>	first generation
<b>2 G</b>	Second Generation
<b>CDMA</b>	Code Division Multiple Access
<b>IS-95</b>	Interim Standard 95
<b>PDC</b>	Personal Digital Cellular
<b>IDEN</b>	Integrated Digital Enhanced Network
<b>EDGE</b>	Enhanced Data Rates for GSM Evolution
<b>GPRS</b>	General Packet Radio Service
<b>3 G</b>	Third Generation
<b>W-CDMA</b>	Wideband Code Division Multiple Access
<b>SMS</b>	Short Message Service
<b>MMS</b>	Multimedia Messaging Service "Service de Messagerie Multimédia "
<b>MS</b>	Station Mobile
<b>BSS</b>	Base Station Subsystem
<b>NSS</b>	Network Sub-System
<b>OSS</b>	Operations Support System
<b>IMEI</b>	International Mobile Equipment Identity
<b>33 db</b>	33 décibels (unité indique couramment le niveau du bruit).
<b>IMSI</b>	International Mobile Subscriber Identity
<b>MSISDN</b>	Mobile Station Integrated Services Digital Network Number
<b>TMSI</b>	Temporal Mobile Subscriber Identity
<b>MSC</b>	Mobile Switching Center
<b>VLR</b>	Visitor Location Register
<b>BTS</b>	Base Transceiver Station (station de base)
<b>BSC</b>	Contrôleur de la Station de Base
<b>AUC</b>	Centre d'Authentification
<b>EIR</b>	Equipment Identity Register

# Acronyme et abréviation

---

<b>IMEI</b>	International Mobile Equipment Identity
<b>NMC</b>	Network Management Center
<b>OMC</b>	Operation and Maintenance Center
<b>UIT</b>	Union Internationale des Télécommunications
<b>IMT-2000</b>	Systems International Mobile Telecommunications-2000
<b>CDMA</b>	Code Division Multiple Access
<b>3GPP</b>	3rd Generation Partnership Project
<b>UE</b>	Equipment usager
<b>MT</b>	Mobile Termination
<b>TE</b>	Terminal Equipment
<b>carte USIM</b>	Universal Subscriber Identity Module.
<b>UTRAN</b>	Universal Terrestrial Radio Access Network
<b>RNS</b>	Radio Network Subsystem
<b>RNC</b>	Radio Network Controller
<b>E Node B</b>	Evolved Node B
<b>FDD</b>	Frequency Division Duplexing
<b>TDD</b>	Time-Division Duplex
<b>TDMA</b>	Time Division Multiple Access
<b>AMRT</b>	Accès Multiple à Répartition dans le Temps
<b>JPDC</b>	Japanese Personal Digital Cellular
<b>TS</b>	Time Slots
<b>IAM</b>	Interférences d'Accès Multiple
<b>FDMA</b>	Frequency Division Multiple Access
<b>AMRF</b>	Accès Multiple par Répartition de Fréquences
<b>TH-CDMA</b>	Time Hopping Code Division Multiple Access
<b>FH-CDMA</b>	Frequency-hopping Code Division Multiple Access
<b>DS-CDMA</b>	Direct Sequence Code Division Multiple Access
<b>DSSS</b>	Direct Sequence Spread Spectrum
<b>BER</b>	Bit Error Rate
<b>S/N</b>	Le rapport de puissance signal/bruit

# Acronyme et abréviation

---

<b>PN</b>	Pseudo-Noise
<b>CRISP-DM</b>	Cross-Industry Standard Process for Data Mining
<b>MATLAB</b>	Matrix laboratory
<b>WEKA</b>	Waikato environment for knowledge analysis,
<b>Eb</b>	Energies de Bits
<b>ARFF</b>	Attribute-Relation File Format

## Introduction Générale :

De nos jours, les plus grands responsables dans tout les secteurs se lient pour donner une plus grande place à la programmation dans les marchés publics. Ensemble, ils cherchent des moyens de la rendre plus accessible et mieux structurée. Première solution, l'utilisation des algorithmes d'apprentissage machine (Machine Learning ou ML) nécessitant des données en grand nombre et bien précises. La structure pourra être laissée entre les mains d'experts. Ensuite, une considération plus sérieuse de la qualité des données et ses multiples usages

Data mining consiste à mettre en œuvre un certain nombre de techniques visant à permettre aux machines d'imiter une forme d'intelligence réelle. Il se retrouve implémentée dans un nombre grandissant de domaines d'application

Dans ce thème de recherche, nous utiliserons cette technologie de Data mining dans notre spécialité afin de garantir une utilisation optimale de la technologie afin de réduire au maximum les erreurs de transmission lie au technique DS-CDMA .

Ce travail de fin d'étude comporte trois chapitres, dont deux théoriques et un chapitre pratique :

Dans le premier chapitre nous allons entamer les notions de base d'un système de communication numérique, nous allons parler plus précisément sur les généralités sur les réseaux cellulaires.

Le deuxième chapitre est consacré aux principes de l'étalement de spectre par séquence directe (DS-SS), plus précisément sur les techniques d'accès multiple. On précise dans cette partie le principe les avantages et inconvénients de chaque technique. En effet c'est la technique CDMA qui nous intéresse le plus puisqu'elle est basée sur l'étalement de spectre.

Dans le troisième et dernier chapitre, Nous avons simulé les paramètres du (DS-SS) pour avoir un bon résultat du signal rapport sur bruit (BER) en utilisée logiciel qu'est très utiliser dans les différent classification et apprentissage de Data mining ce logiciel s'appelle "Weka" .

---

# *Chapitre 1 :*

## *Généralités sur les Réseaux cellulaires*

## I.1. Introduction

Le GSM et l'UMTS sont des réseaux de communications cellulaires, leur apparition est considérée comme un grand pas dans l'évolution des réseaux mobiles, grâce aux nouveaux services qui ils ont mis à la disposition de tous les utilisateurs avec un débit satisfaisant.

Dans ce chapitre nous allons donner une vue globale de ces systèmes, en commençant tout d'abord de définir le concept cellulaire puis en abordant leurs principes de fonctionnement, l'architecture de ses différentes Entités et quelques fonctions avancées sans rentrer dans le détail.

## I.2. Concept cellulaire

La transmission sans fil consiste à définir une zone de couverture et installer une antenne relais qui servait d'un point d'accès aux utilisateurs évoluant dans cette zone, la puissance d'émission doit être importante et capable d'atteindre la périphérie de la zone de couverture mais l'atténuation du signal limite cette zone en termes de puissance et de capacité, ce qui a permis de développer un concept cellulaire.

Le principe consiste à diviser une région en un certain nombre de cellules, la cellule est un modèle théorique représenté par un hexagone et qui ressemble le plus au cercle, la juxtaposition des cellules permet de couvrir une zone géographique sans laisser d'espaces vides [1].

### I.2.1. Type des cellules :

La taille d'une cellule varie en fonction d'un ensemble de contraintes telle que : la topologie du terrain, la densité d'abonnés et la nature de l'environnement, c'est ainsi que les opérateurs ont recours à ces 3 types de cellules :

- **Macro cellules** : sont des cellules dont leurs zones d'action s'étendent jusqu'à 30 km, desservant un trafic moins élevé.
- **Micro cellules** : sa portée maximale est de 500 m environ, couvrant ainsi des zones urbaines est desservant un trafic moyennement dense
- **Pico cellules** : ces pico cellule sont placée dans des endroits où le trafic est considérable.





Figure 1.1 : types de cellules

## 1.3. Evolution des réseaux cellulaires :

Les réseaux mobiles ont connu un succès phénoménal lors de leur déploiement, ils proposent une meilleure mobilité et une qualité de service à la rigueur par rapport aux réseaux classiques, tels que la téléphonie fixe. Depuis leur création et leur mise en service les réseaux mobiles ont marqué une forte adoption auprès des professionnels et de grand public, ainsi que l'intégration des services et la diversité des données échangées exigent des systèmes de plus en plus développés [2].

### 1.3.1. La première génération (1G) :

Le début des années 1980 fut marqué par l'ouverture du réseau radio analogique de 1<sup>ère</sup> génération (Radiocom2000 en France). Celui-ci offrait un service de radio téléphonie mobile avec des terminaux lourds, encombrants et chers. Le temps de connexion vers les abonnés filaires était long, ces réseaux n'étaient pas universels.

### 1.3.2. La deuxième génération 2G (GSM) :

La deuxième génération de réseaux mobiles est apparue comme étant une amélioration de la première génération. À l'instar du GSM (Global system for mobile communications) qui est apparu en Europe en 1991 et qui a connu un succès très vif, d'autres systèmes existent encore comme le IS-95 (ou CDMAOne) en Amérique du nord, PDC en Japon, IS-136 et IDEN en U.S.A.

Le tableau suivant résume les spécificités techniques de chaque standard :

Standards	GSM	IS-95	IS-136	PDC
modulation	GMSK	BPSK	DQPSK	QPSK
Méthode d'accès	TDMA / FDMA	CDMA	TDMA	TDMA
Bande de fréquence (MHz)	900/1800/1900	800/1900	800/1900	800/1400
Espacement	200	1250	30	25
Utilisation	Mondial	Continent Américain, Asie	Amérique du nord	Japon

**Tableau 1.1 : principales différences technologiques entre le GSM, IS-95, IS-136 et le PDC**

### 1.3.3. Evolution vers l'EDGE :

L'EDGE (Enhanced Data Rates for GSM Evolution) appelé aussi 2.75G est une norme de la téléphonie mobile qui permet un débit triple que celui de GPRS et une capacité plus importante.

### 1.3.4. La troisième génération (3G) :

La troisième génération est marquée par l'apparition de nouveaux systèmes de communications radio mobile tel que l'UMTS connu aussi sous le nom W-CDMA, ces systèmes sont une amélioration des spécificités techniques de la norme européenne GSM, en effet les principaux facteurs qui ont permis la conception et la mise en œuvre de ce nouveau modèle universel sont :

- Les technologies différentes
- Incompatibilité des systèmes
- La possibilité de rester connecté en dehors de la zone d'abonnement (roaming).

## I.4. Présentation du réseau GSM :

Est l'une des normes de la deuxième génération des réseaux mobiles cellulaires apparue en Europe dans les années 90, cette nouvelle technologie permet des communications numériques entre les abonnés mobiles avec une qualité de service meilleur et une confidentialité approuvée. Les services que propose la norme GSM sont la transmission de la voix ainsi que des données numériques de faible volume comme les SMS et MMS [3] .

**I.4.1. Architecture du réseau GSM :** Le réseau GSM est constitué de quatre sous-systèmes :

- La station mobile (MS)
- Le sous -système radio (BSS)
- Le sous-système réseau (NNS)
- Le sous-système d'exploitation et de maintenance (OSS)

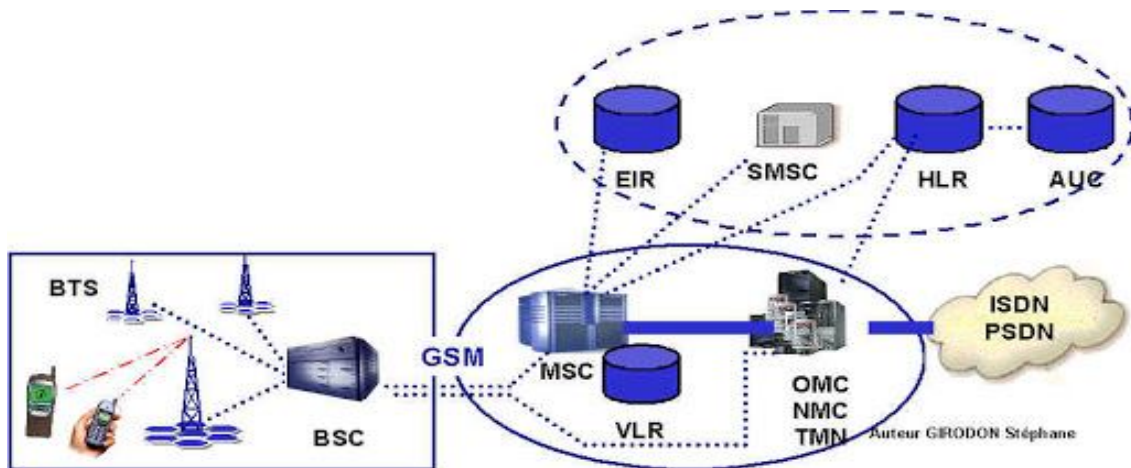


Figure 1.2 : schéma en bloc des différentes entités du réseau GSM

### 1.4.1.1. Station mobile (MS) :

Est le terminal mobile authentifié et autorisé à accéder au réseau mobile, il se compose de deux entités à savoir:

- **Equipment terminal (Mobile Equipment)** : il s'agit de l'équipement mobile doté à la base d'un système émetteur/récepteur radio mobile et d'un logiciel nécessaire à la communication. Il contient une identité unique IMEI (International Mobile Equipment Identity), il est caractérisé par une puissance d'émission maximale (2W soit 33dB pour le GSM 900)
- **La carte SIM** : est la partie personnalisée du terminal qui différencie un abonné des autres, en contenant l'identité de l'abonné et les caractéristiques de l'abonnement tel que :
  - IMSI (International Mobile Subscriber Identity) : unique et universel
  - MSISDN : numéro d'appel
  - TMSI : (Temporal Mobile Subscriber Identity)

### 1.4.1.2. Le sous-système radio (BSS) :

Il assure l'accès des abonnés au réseau GSM avec l'interface radioélectrique d'une part et la liaison câblée avec les MSC /VLR d'une autre part, il comporte deux types d'équipements à savoir les BTS et les BSC.

## La station de base (BTS) :

La BTS est l'élément central du BSS comporte un ensemble d'émetteurs (Tx) et récepteurs (Rx), servant d'interface entre le réseau fixe et les stations mobiles. Elle se charge de bon déroulement de la transmission radio en réalisant des relevés de mesures instantanées qu'elle envoie directement au BSC sans les interpréter. Il existe deux types de BTS les macros et les micros BTS, contrairement aux premières qui se placent dans des régions rurales et qui couvrent des surfaces plus étendues, ces dernières sont utilisées pour couvrir des zones urbaines avec un trafic considérable, elles sont caractérisées par leurs faibles tailles, faible cout et moindre puissance.

## Le contrôleur de la station de base (BSC) :

Considéré comme l'organe le plus intelligent du réseau, il gère une ou plusieurs stations de bases en assurant le contrôle de la puissance de signal, l'acheminement du trafic vers les stations de bases destinataires, il est aussi le lien entre la BTS et le NSS, Que nous verrons juste après.

### 1.4.1.3. Le sous-système réseau (NSS) :

C'est la partie invisible du réseau, elle assure les fonctions d'acheminement et de routage des données, la gestion de la mobilité des abonnés, elle enregistre les informations propres à chaque abonné. Contient les organes suivants :

#### ➤ MSC:

Est un commutateur chargé de raccorder plusieurs BSC, raccordée aussi à d'autres MSC existant dans le réseau réalisant ainsi un maillage des MSC, il a comme fonctions principales :

- Réserve de canal radio.
- Acheminement vers le demandé.
- Taxation des appels et les services de données.
- Localisation des usagers.
- Gestion de la mobilité.

Le MSC communique aussi avec les deux bases de données le HLR et VLR.

#### ➤ GMSC :

C'est la passerelle qui gère les communications avec les autres réseaux externes tels que le réseau téléphonique les autres réseaux GSM et les réseaux de transmission de données.

➤ La Base de données HLR :

Est un registre de données qui contient toutes les informations liées aux abonnés du réseau GSM, Ces informations sont :

- Le Profil de chaque usager
- Le numéro IMSI
- Le numéro d'appel de l'abonné
- Le numéro de VLR où l'utilisateur est localisé

➤ La base de données VLR :

Est un registre de données temporaire contenant les données de travail relatives aux abonnés présents dans la zone de localisation gérée par un MSC, le réseau a prévu l'installation de cet organe dans le but de minimiser les accès au HLR.

➤ Le centre d'authentification (AUC) :

Est l'élément important associé au HLR dans le but d'assurer la sécurité des appels en effectuant le chiffrement des échanges ainsi que l'authentification des abonnés.

➤ La base de données (EIR) :

Est un registre d'identification d'équipements, installé dans le but d'identifier tous les mobiles volés ou interdits d'accéder au réseau, regroupés dans une liste appelée black liste. A chaque établissement d'appel la phase d'authentification des terminaux se fera en premier lieu par la demande de IMEI, si ce dernier retourné par le terminal fait partie de black liste, l'appel ne peut être établi.

#### 1.4.1.4. Le Sous System d'Exploitation et de Maintenance (OSS) :

Il a comme fonction le contrôle et la gestion de réseau en détectant d'éventuelles pannes, mettre en services de nouveaux sites, observation de trafic et détection des surcharges, configuration des équipements et modifications de leurs paramètres, il est subdivisé en deux parties :

**NMC** : il opère d'une manière centralisée (générale)

**OMC** : il effectue le contrôle local des équipements

## 1.5. Présentation d'un réseau UMTS :

On a vu précédemment que les réseaux mobiles de la 2eme génération ont connus une multitude de systèmes et de technologies différentes, dont chaque technologie appartienne à un organisme de normalisation, chargé du suivi de l'évolution de ses spécifications

Afin de ne pas laisser qu'un tel scenario ne se reproduise pour la 3eme génération et pour éviter les problèmes qu'engendre la particularité que connait chaque système de la 2eme génération, il devenait très nécessaire de placer la définition d'un tel système sous la responsabilité d'un organisme aussi indépendant et fédérateur que possible.

De la part des opérateurs et constructeurs de technologies radio, il y a eu une forte volonté de définir une norme commune, c'est ainsi que l'UIT a nommé le IMT-2000, visant à tenir compte des occupations que manifeste chaque organisme de normalisation et de parvenir à une norme commune universel ayant pour objectifs :

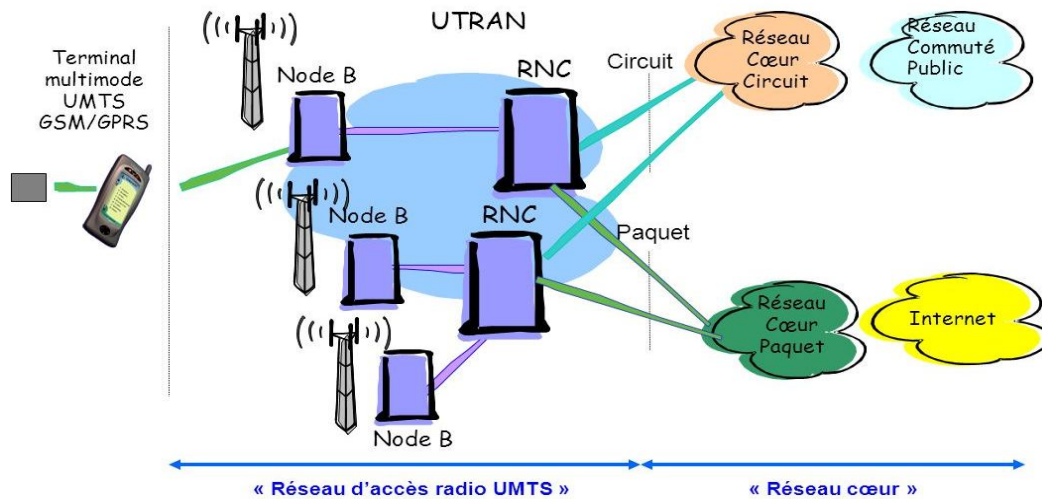
- Une compatibilité avec les systèmes de la deuxième génération.
- Un support des applications multimédia.
- Un support débit très élevé.

Lors de la soumission des propositions techniques provenant de différents organismes, deux types ont été constatées à savoir les propositions pour des réseaux de communications terrestres et celles des constellations de satellites d'où la technologie CDMA était majoritairement gagnante. La phase de négociation a été marquée par la création des deux regroupements de constructeurs et d'opérateurs de réseaux :

- Le 3GPP : à l'origine de l'UMTS, dont la définition a hérité d'un certain nombre de concepts de la norme GSM
- Le 3GPP2 : crée en réaction à la forte inclinaison au GSM du 3GPP, afin de garantir une certaine pérennité des systèmes de types IS-95, concurrent nord-américain du GSM.

### I.5.1. L'architecture du réseau UMTS :

Comme nous l'avons vu précédemment, l'architecture d'UMTS est similaire à celle du réseau GSM, qui se subdivise en trois entités essentielles, comme il avait spécifié le groupe de normalisation 3GPP, d'où il comprend l'unité terminal, un réseau d'accès et un réseau cœur.



**Figure 1.3: Architecture globale du réseau UMTS**

### I.5.1.1. L'équipement usager UE :

L'apparition de L'UMTS est accompagnée de l'apparition de nouveaux modèles de terminaux 3G, qui ont connus une certaine évolution par rapport aux précédents en termes d'applications et supports multimédia que propose le réseau UMTS, composé de deux blocs suivants:

- **L'équipement mobile ME** : qui est un terminal radio, découpé à son tour en deux entités:
  - Le MT (Mobile Termination) : il se charge de la transmission radio.
  - Le TE (Terminal Equipment) : il regroupe les services et applications qui sont accessibles directement par l'abonné comme le navigateur Web.
- **La carte USIM** : est l'entité personnalisé de réseau, elle joue le même rôle que la carte SIM des terminaux GSM.

### I.5.1.2. Le réseau d'accès UTRAN :

L'UTRAN est la partie de réseau UMTS qui s'interface entre la station mobile et le réseau cœur, qui possède les fonctions principales suivantes : sécurité, mobilité, gestion des ressources radio et la synchronisation.

L'UTRAN se compose du plusieurs RNS qui sont des sous -systèmes réseau radio, dont chaque RNS est constitué d'un RNC, équivalent du BSC des réseaux GSM, ils servent à contrôler un ensemble du Node B , Équivalent de la BTS, et des interfaces de communication entre les différents éléments du réseau UMTS.

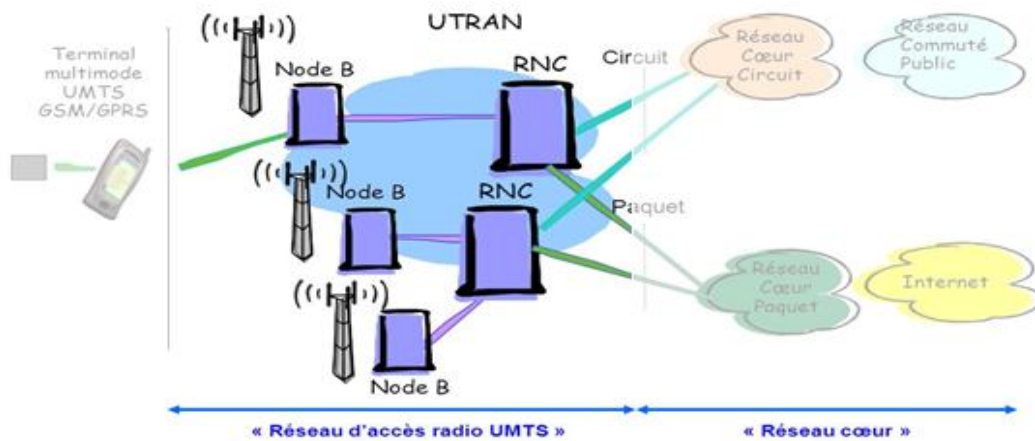


Figure 1.4 : Architecture matérielle de réseau UTRAN du réseau UMTS

**Node B** : Son rôle principal est d'assurer les fonctions de transmission et de réception radio pour une ou plusieurs cellules de l'UTRAN.

**Le RNC** : Est l'équivalent de BSC des réseaux GSM, son rôle principal est le routage des communications entre la station de base (Node B) et le réseau cœur d'une part, et de contrôler plusieurs Node B d'autre part .



**I.5.2. Organisation des fréquences dans un réseau UMTS :**

La norme UMTS a intégré la possibilité d'utiliser les ressources radio en deux modes différents FDD et TDD C'est-à-dire les fréquences sont partagées en deux :

- TDD : deux plages ont été occupées par le mode TDD, dont chacun est de 35 MHz
- FDD : il occupe une plage de 120 MHz de spectre des fréquences de réseau UMTS, la communication dans le sens montante (mobile vers Node B) est supportée par une bande de fréquence de 60 MHz [de 1920 à 1980 MHz] disjointe de celle utilisé pour la voie descendante (BTS vers mobile) qui est de 60 MHz aussi [de 2110 à 2170], L'écarte entre les deux bandes est appelé écart duplex qui est de 190 MHz.

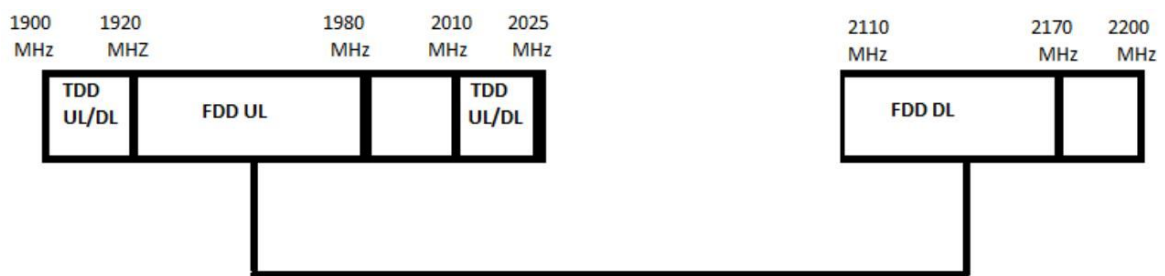


Figure 1.5 : la répartition des fréquences d'un réseau UMTS

**I.5.3. Organisation temporelle :**

Le système UMTS repose sur une technique de multiplexage par code dite CDMA qui est organisée à partir d'une super-trame de 720ms l'équivalente de 72 trames d'UMTS dite trame de base .

**Trame de base :** est la trame temporelle de 10 ms, elle se compose de 15 intervalles de temps dite Slots dont chacun est de 10/15 ms équivalente de 666.67 micro s .la communication entre une station mobile et un usager quelconque se fait sur tous les 15 slots successifs de la trame UMTS.

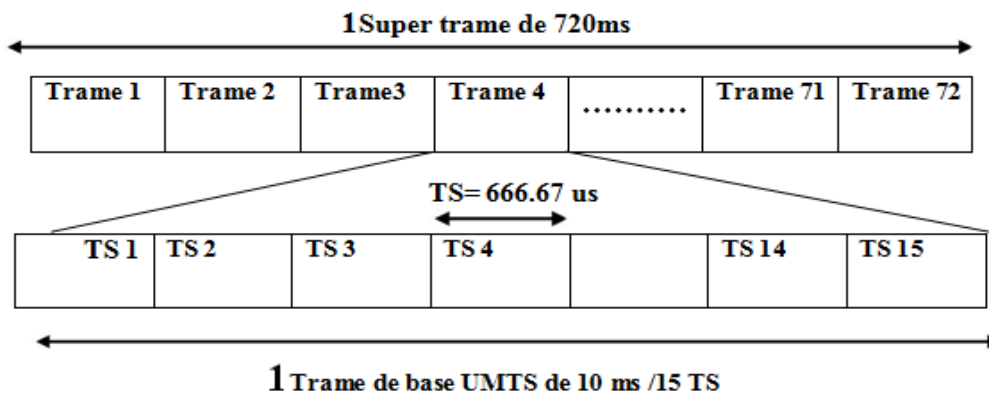


Figure 1.6 : la trame temporelle dans un système UMTS

## I.5.4. La technique d'accès dans un réseau UMTS :

Dans le but d'autoriser à un ensemble d'utilisateurs d'accéder au même temps au réseau, il est nécessaire de partager la bande de fréquence allouée par l'opérateur, la manière dont ce partage se fait, peut-être regroupé dans trois grandes techniques appelées techniques d'accès multiple :

- Le partage en fréquences FDMA
- Le partage en temps TDMA
- Le partage en code CDMA

### 1.5.4.1 . Accès multiple par répartition dans le temps ou "Time Division Multiple Access (TDMA)" :

Beaucoup de systèmes de communications numériques comme le "Global System for Mobiles communications " (GSM) européen, utilisent l'Accès Multiple à Répartition dans le Temps (AMRT), aussi appelé Time Division Multiple Access (TDMA). Cette technologie est aussi incluse dans le système de communication cellulaire japonais (Japanese Personal Digital Cellular : (JPDC) ) qui représente le second standard le plus utilisé [4] .

Le multiplexage en temps consiste à répartir la transmission d'une porteuse sur des intervalles de temps distincts, comme l'illustre la Figure 1.7. la porteuse est émise sur des intervalles de temps, appelés time slots de durée  $T_s$  , chacune de ces durées correspondant à un temps pendant lequel un utilisateur peut transmettre ses données .

Un émetteur / récepteur, tel qu'un téléphone cellulaire, peut, par exemple, émettre et recevoir dans un intervalle de temps spécifique et rester inactif jusqu'à ce qu'il soit autorisé à intervenir de nouveau. En d'autres termes, chaque intervalle de temps accueille une partie du signal radiofréquence associé à un utilisateur donné. La porteuse du signal est divisée en plusieurs séquences appelées "Burst", sachant qu'un Burst est envoyé par intervalle de temps. Il en résulte que les données émises peuvent être transmises à la même fréquence sans entraîner d'Interférences d'Accès Multiple (IAM).

L'allocation des intervalles de temps aux différents utilisateurs se fait de manière cyclique Comme le montre la Figure 1.7. Il est donc possible de caractériser un système TDMA par la durée d'un intervalle de temps  $T_s$  , la durée d'un cycle ( $T_{TDMA}$  ) et le numéro de l'intervalle de temps associé à chaque utilisateur ( $N_T$  )

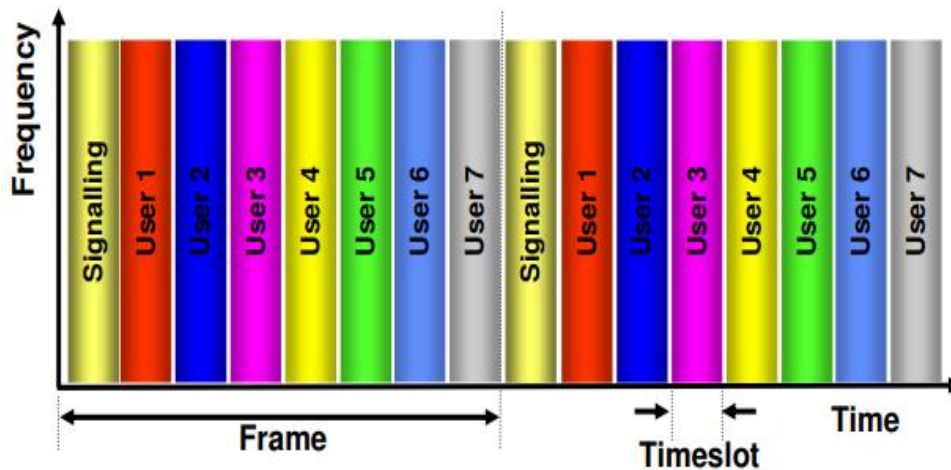


Figure 1.7 : Schéma d'une multiplexage temporel (TDMA)

Le nombre d'intervalles de temps, la taille du cycle, et la largeur de bande employée dépendent de la technologie utilisée. Dans le cas de la norme GSM, par exemple, la durée élémentaire d'une intervalle de temps, sur une horloge de 13 MHz, vaut :

$$T_s = (75/130) \times 10^{-3} = 0.5769 \text{ ms}$$

Sur une même porteuse, le time slots sont regroupés par paquets de 8, ce qui implique que la durée d'une trame TDMA est donnée par :

$$T_{TDMA} = 8 \times T_s = 4.6152 \text{ ms}$$

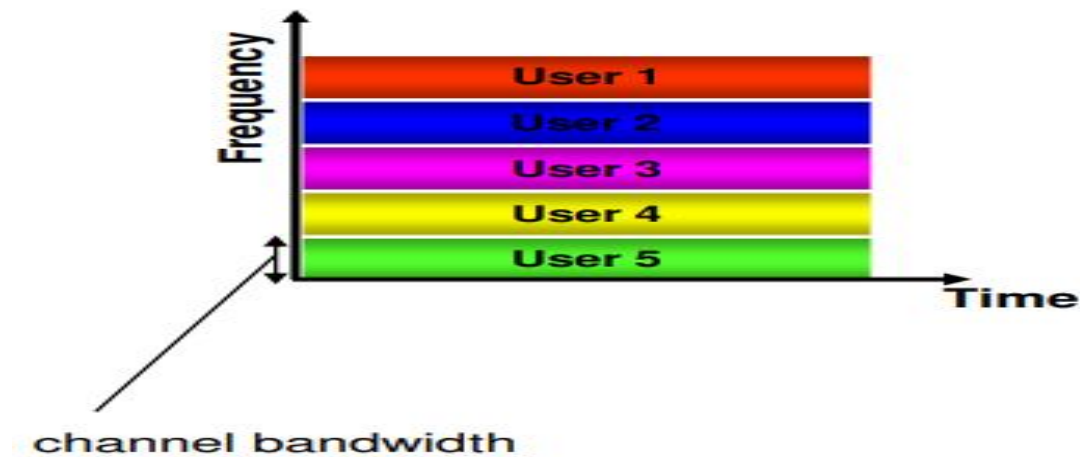
A chaque usager est affecté un certain nombre d'intervalles par trame TDMA, numérotés par un indice TN variant entre 0 et 7, Un canal physique est, par conséquent, constitué par la répétition périodique d'un intervalle dans la trame TDMA sur une fréquence donnée .

La majorité des systèmes de communications radiofréquences utilisant le TDMA sont dit synchrones, ce qui implique une gestion des périodes d'émission de chacun des utilisateurs .

Afin de ne pas engendrer d'IAM, il est impératif que chaque signal respecte strictement l'intervalle de temps qui lui est accordé pour sa transmission. Les récepteurs doivent aussi respecter les intervalles de temps qui leurs sont associés pour reconstituer correctement les séquences d'information qui leurs sont destinées , Pour ce faire au début de chaque transmission, une séquence de synchronisation est envoyée afin que l'émetteur et le récepteur puissent identifier le numéro et la durée du time slot sur lequel ils pourront échanger leurs informations , L'un des avantages de cette technique est de s'affranchir de la gestion des fréquences des séquences de données à transmettre .

## 1.5.4.2. Accès multiple par répartition de fréquences ou "Frequency Division Multiple Access (FDMA)" :

L'Accès Multiple par Répartition de Fréquences (AMRF), appelé aussi Frequency Division Multiple Access (FDMA), est une technique de multiplexage largement utilisée dans les systèmes de communications radiofréquences. Le FDMA consiste à transmettre les signaux provenant de N utilisateurs différents sur, des bandes de fréquences distinctes comme il est montré dans la Figure 1.8 qui représente l'accès multiple de cinq utilisateurs avec la technique FDMA [4] :



**Figure 1.8 : Schéma d'un multiplexage fréquentiel (FDMA)**

Pour plus de compréhension en illustrant la Figure 1.9 qui donne un exemple de la manière dont ce multiplexage s'effectue :

- **Au niveau d'émetteur :** d'abord il existe un filtre passe bas est utilisé afin de s'assurer que la bande passante du message à transmettre est limitée à une largeur définie au préalable. Chaque signal module une porteuse différente dont la fréquence est fixée par le synthétiseur. Les signaux issus des N modulateurs sont superposés et transmis par le canal.
- **Au niveau de récepteur :** les porteuses sont séparées par des filtres passe bande placés en parallèle et centrés sur les fréquences porteuses des messages à détecter. La largeur de ces filtres est déterminée afin de ne pas tronquer le spectre du signal. A la sortie du filtre passe bande, le signal est démodulé et traverse un filtre passe-bas permettant d'écartier les composantes indésirables et de récupérer le message en bande de base.

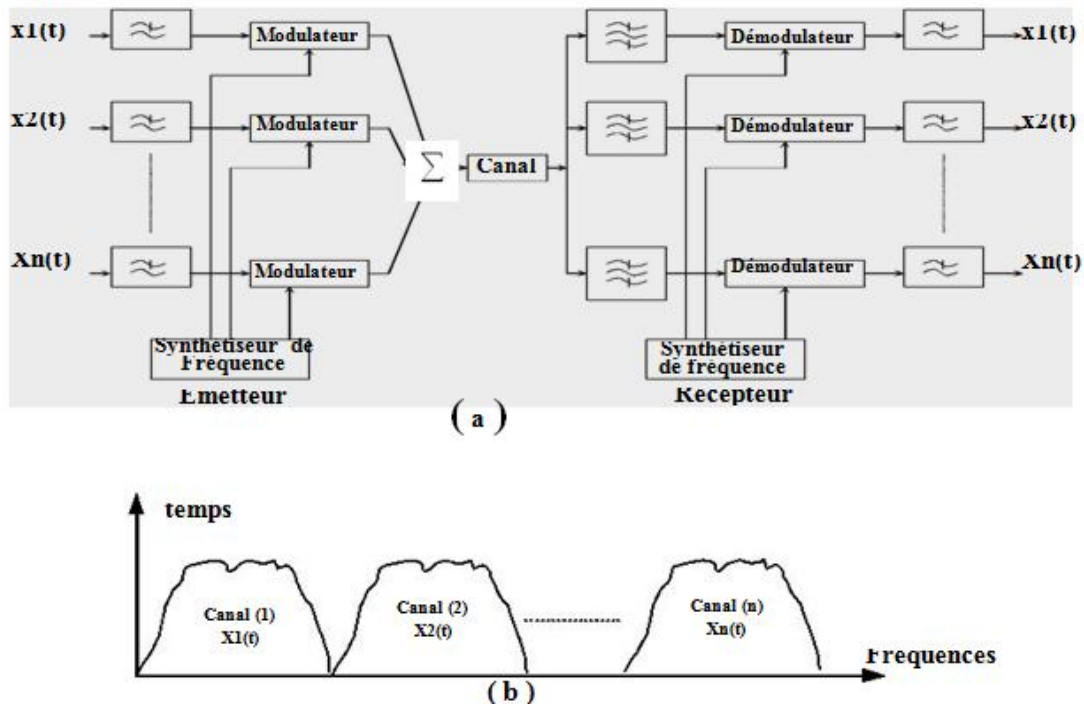


Figure 1.9 : Principe de multiplexage en fréquence.

Dans ces conditions, les utilisateurs émettent leurs signaux de manière continue, chacun sur une fréquence différente. La suppression des interférences est assurée par l'utilisation de fréquences porteuses différentes et séparées

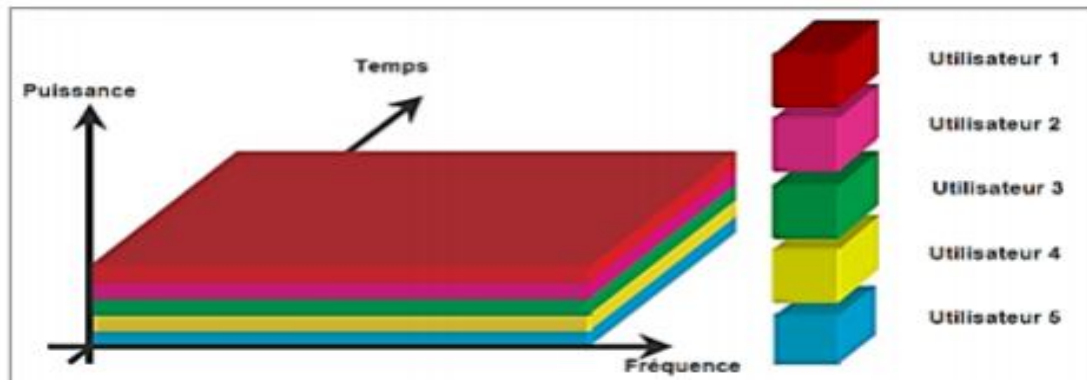
par un intervalle fréquentiel prédéfini représentant la largeur du canal affecté à un utilisateur (Figure 1.9.b)

### 1.5.4.3. Accès multiple par répartition de code ou "Code Division Multiple Access (CDMA)" :

Le CDMA ou en anglais (Code Division Multiple Access) est une méthode d'accès multiple à un médium de communication multiple à un médium de communication par répartition de code : plusieurs usagers ont alors accès à un canal commun et peuvent l'utiliser simultanément jusqu'à une certaine limite d'utilisateurs actifs définie par la tolérance, la capacité du système. C'est une technologie d'étalement du spectre qui est utilisée depuis longtemps par les militaires pour sa résistance à l'interférence et pour le niveau de sécurité qu'elle offre. C'est une technique qui consiste à redistribuer et étaler le signal sur une très grande largeur de bande, jusqu'à le rendre « invisible » idéalement, pour les autres utilisateurs de la même largeur de bande. Au récepteur, l'opération d'étalement exécutée au transmetteur est répétée pour désétalement le signal en bande de base (ou une fréquence intermédiaire) tandis que les autres signaux transmis (interférence) sont perçus par le récepteur comme étant du bruit [4].

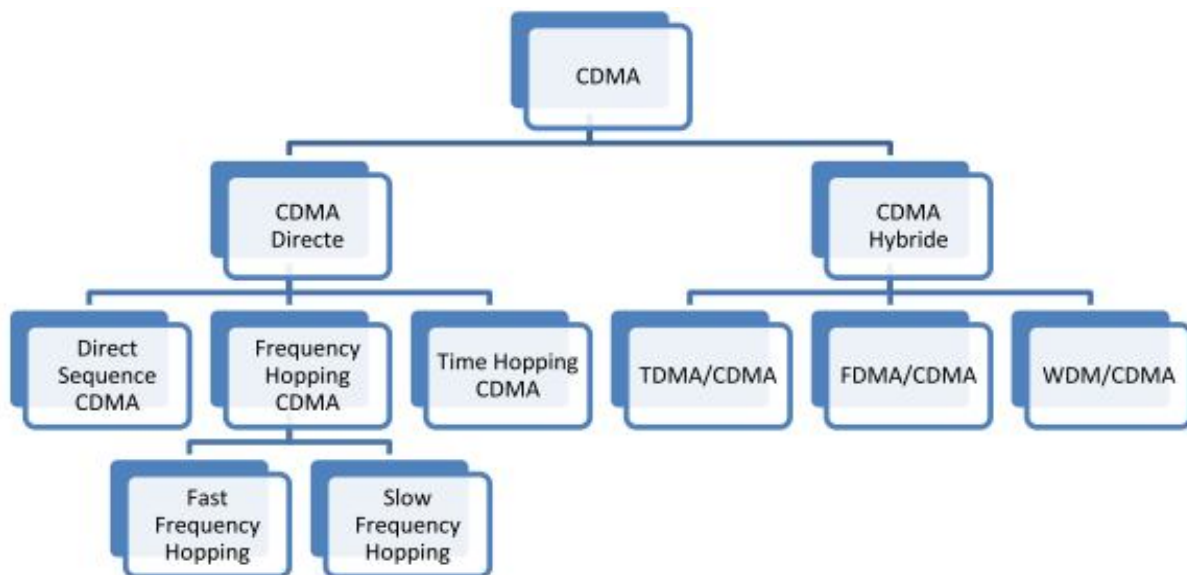
Dans cette technique d'accès, les utilisateurs partagent le même espace fréquentiel et transmettent sur les mêmes intervalles temporels. Il s'agit dans ce cas, d'affecter à chaque émetteur un code, aussi appelé signature qui lui permet de transmettre des informations en évitant d'interférer avec les messages provenant d'autres utilisateurs. En autre façon la technique CDMA permet la transmission simultanée des données aux utilisateurs

sur la même bande de fréquences (figure 1.10). De ce fait, tous les utilisateurs ont accès à la totalité de la bande et ils sont distingués à la réception par des codes distincts qui leurs sont propres .



**Figure 1.10** Technique d'accès multiple CDMA.

Dans la technique d'accès multiple CDMA , les méthodes permettant l'étalement de spectre sont regroupées en deux catégories comme le montre la Figure 1.11, La première appelée CDMA hybride ,regroupe les méthodes qui combinent le CDMA avec les autres techniques de multiplexages (TDMA, FDMA,...). La deuxième c'est la CDMA directe, regroupe TH-CDMA, FH-CDMA et la DS-CDMA. La méthode TH-CDMA consiste à transmettre les chips qui composent une séquence de code donnée sur différents time slots, dans la CDMA à saut de fréquence (FH-CDMA), la fréquence du signal modulé change périodiquement pendant des intervalles de temps et dans la technique DS-CDMA, le signal d'un utilisateur directement en le multipliant par un code d'étalement.



**Figure 1.11** : Différent technique CDMA

Dans ce qui suit nous nous intéressons seulement à l'étalement de spectre par séquence direct, c'est à dire, la technique DS-CDMA [3].

## 1.6. Conclusion :

Ce premier chapitre nous a permis dans un premier temps de rappeler les notions de base d'un système de communication numérique, ensuite les différents systèmes d'accès multiple qui sont derrière l'évolution des communications sans fil. La première le FDMA alloue à chaque usager une bande de fréquence pour communiquer sur tout l'intervalle du temps. La seconde TDMA divise l'intervalle de temps permettant à chaque usager d'utiliser toute la largeur de la bande de fréquence sur son intervalle de temps. La troisième CDMA permet à plusieurs utilisateurs l'échange d'informations simultanément sur la même plage de fréquence et le même intervalle temporel en attribuant à chaque utilisateur un code qui lui est propre. La dernière (Space Division Multiple Access) est basée sur l'utilisation des antennes directionnelles, chaque antenne est orientée selon un angle spécifique. Cette technique est généralement combinée avec les autres techniques d'accès. Dans le chapitre suivant nous allons présenter la technique d'étalement de spectre qui est la base de CDMA.

# *Chapitre 2:*

*Principe de l'étalement de spectre  
par séquence directe 'DS-CDMA'*



## **2.1. Introduction :**

Les systèmes de téléphonie mobile de la troisième génération (3G) utilisent la technique d'accès multiples CDMA basée sur l'étalement de spectre. Cette technique offre une solution plus flexible, par rapport à celles des deux techniques TDMA et FDMA, et surtout un débit utilisateur beaucoup plus important, due à la largeur de bande allouée au signal émis, permettant ainsi des services multimédia très attractifs. En plus, l'étalement de spectre possède des qualités très avantageuses, telles que la résistance au brouillage intentionnel et surtout une parfaite protection contre l'interception de la communication par des intrus. C'est pour toutes ces raisons que l'usage initial de cette technique fut très attractif.

## **2.2. L'étalement de spectre par séquence directe :**

De conception plus récente, ce procédé vise particulièrement les applications civiles radio mobile de type cellulaire pour communications personnelles, en association avec une interface radioélectrique de type CDMA d'où l'appellation (systèmes DSSS-CDMA). Dans son principe, la techniques DSSS vise à réduire la densité spectrale de puissance du signal à émettre en l'étalant sur une bande de fréquence de très grande largeur. Le procédé DSSS de modulation à étalement de spectre constitue une technique d'accès particulièrement souple. Il permet, entre autre [5] :

- De transmettre simultanément des signaux à bande étroite émis par divers utilisateurs dans la même bande de fréquence et sans coordination.
- De réaliser un système de communication à accès multiples à répartition par le code (CDMA). Chaque utilisateur communique via une signature (code personnel) qui permet de distinguer son signal de ceux des autres personnes.
- De partager contre l'interférence intensionnelle (brouillage).
- De réduire la possibilité de détecter le signal en le cachant dans le bruit de fond.
- D'assurer, par un traitement original, une protection contre les interférences aléatoires dues aux trajets multiples engendrés par la propagation. Ce dernier point engendre un intérêt tout particulier dans le domaine radio mobile [6].

Notons que cette dernière est la plus utilisée dans les transmissions de type CDMA. Dans ce cas, on parle de transmission DS-CDMA, dont le principe sera détaillé dans les sections suivantes.

### 2.3 Pourquoi étaler le spectre ?

Considérons le théorème de Shannon et Hartley concernant la capacité d'un canal de communication :

$$C = B \log_2 \left( 1 + \frac{S}{N} \right) \quad (\text{Eq. 2.1})$$

Dans cette équation,  $C$  représente la capacité maximale d'un canal en bits par seconde (bit/s ou bps), c'est le taux de transfert maximum pour un taux d'erreur binaire (Bit Error Rate, BER) nul, à condition qu'un procédé de codage adéquat puisse être trouvé.  $B$  étant la bande passante du canal en Hertz et  $(S/N)$  le rapport de puissance signal/bruit.

On peut donc augmenter la capacité maximale en agissant sur la largeur de bande de façon linéaire et/ou en agissant sur le rapport signal bruit de façon logarithmique.

A capacité maximale donnée (capacité maximale souhaitée), on peut réduire la bande et/ou diminuer le rapport signal à bruit en admettant un taux d'erreur non nul. Les erreurs peuvent être soit tout simplement ignorées soit corrigées par l'utilisation de protocoles de transmission de niveau supérieur. Au niveau de la formule, en fonction du type de bruit et du procédé de codage/décodage, on peut intégrer le BER sous la forme de l'addition d'une constante au rapport signal sur bruit [6].

Dans le cas du CDMA, le bruit est constitué principalement par les autres utilisateurs dont on cherchera à augmenter le nombre. Il en résulte qu'en règle générale un système CDMA opère sur des rapports signal à bruit faibles, voire très faibles. Par changement de base des logarithmes (base 2 vers  $e$ ), l'équation (2.1) devient :

$$\frac{C}{B} = \frac{1}{\ln(2)} \ln \left( 1 + \frac{S}{N} \right) = 1.443 \ln \left( 1 + \frac{S}{N} \right) \quad (\text{Eq. 2.2})$$

Si la puissance du signal est inférieure à la puissance du bruit, on peut simplifier et linéariser l'expression (2.1), en appliquant le développement en série de MacLaurin de  $\ln(1+x)$  :

$$\frac{C}{B} = 1.443 \left[ \frac{S}{N} - \frac{1}{2} \left( \frac{S}{N} \right)^2 + \frac{1}{3} \left( \frac{S}{N} \right)^3 - \dots \right] \quad (\text{Eq. 2.3})$$

Puisque l'étalement du spectre permet un rapport  $(S/N)$  très faible et que la puissance du signal utile pouvant être inférieure au niveau du bruit. Pour un  $(S/N) \ll 1$ , l'équation (2.1) devient alors :

$$\frac{C}{B} = 1.443 \left( \frac{S}{N} \right) \quad (\text{Eq. 2.4})$$

Et par approximation on obtient la relation suivante :

$$\frac{C}{B} \approx \frac{S}{N} \text{ ou } \frac{N}{S} \approx \frac{B}{C} \quad (\text{Eq. 2.5})$$

La dépendance capacité/rapport signal à bruit est approximativement linéaire. La bande étalée permet donc la transmission de signaux perturbés par d'autres signaux considérés alors comme du bruit, c'est-à-dire la transmission de signaux sur le même support. Le nombre de canaux utilisés à un instant donné pourra varier de façon souple puisque l'augmentation du nombre d'utilisateurs se traduira simplement par une augmentation, pour tous, du taux d'erreur. Ceci permet en téléphonie de maintenir une qualité de service sensiblement égale pour tous, (plutôt qu'une dépréciation totale pour un utilisateur) ajustable et relativement facile [5].

### 2.3.1 Exemple d'application :

On décrit ci-dessous un exemple d'étalement par séquence directe (Direct Sequence Spread Spectrum) figure (2. 1).

- Le message **A** de l'émetteur **A**, représentée par une séquence de +1, -1 traduisant la séquence de bits 1 et 0 logiques, est multiplié par un code **CA** d'une séquence de +1 et -1 de chips judicieusement choisie, et dont les transitions sont m fois plus fréquentes. Un autre message **B** de l'émetteur **B** multiplié par un code **CB**.
- Les séquences produits **A\*CA** et **B\*CB** sont ajoutées et transmises.
- A la réception, le destinataire du message **A** multiplie la séquence reçue par le code **CA**, la même opération pour le destinataire du message **B**. Figure (2.1).

Si les codes sont bien choisis, sur la durée d'un bit, (donc de m chips), la moyenne de **CA.CA** et de **CB.CB** est égale à m/2, tandis que **CA.CB** a une moyenne nulle : Les codes **CA** et **CB** sont dits orthogonaux, (produit scalaire=0).

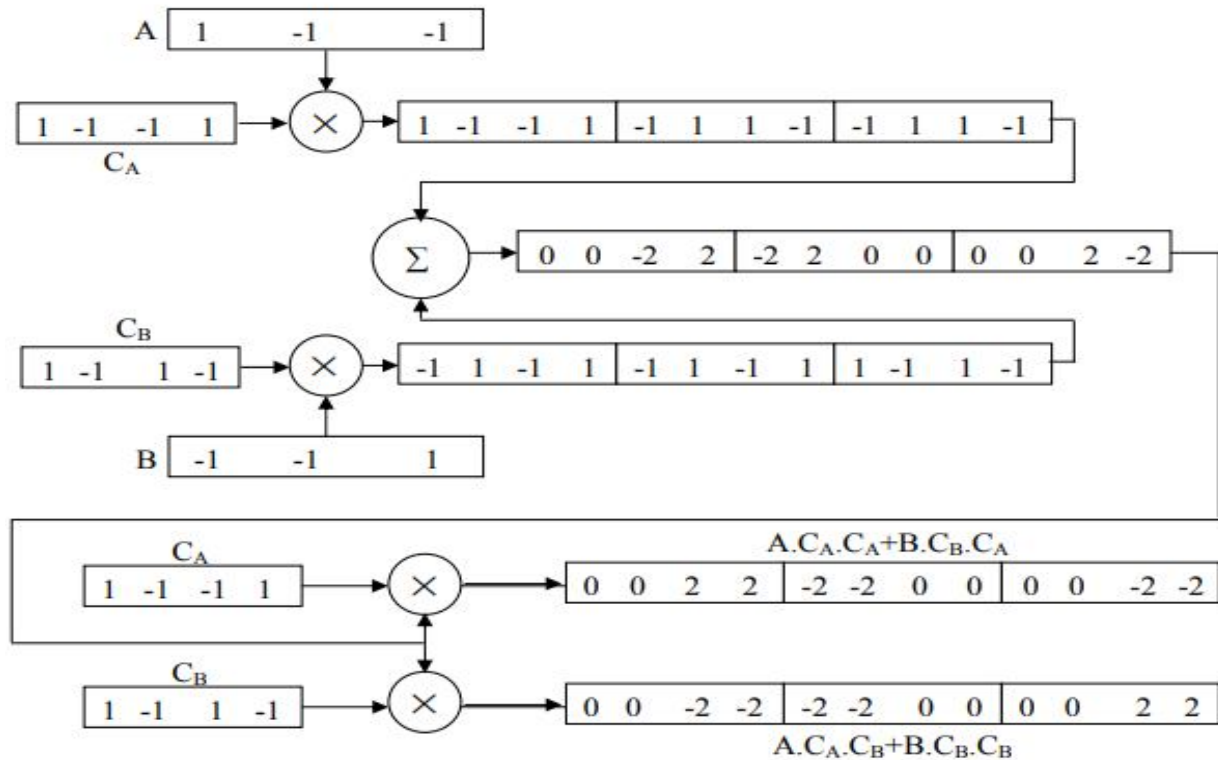


Figure 2.1 : Exemple d'étalement par séquence directe

Remarque :

- La séquence somme est transmise sur trois niveaux d'amplitudes avec deux émetteurs. Quatre avec trois émetteurs etc. La moyenne sur chaque durée d'un bit est nulle.
- Lorsque le nombre d'émetteurs est plus important, la distribution des amplitudes s'apparente à une distribution Gaussienne (comme le bruit Gaussien).
- En réception les signaux sur chaque durée d'un bit de message ont une moyenne non nulle, ce qui permet la reconstitution du signal par simple filtre passe-bas. On préfère en fait mesurer la corrélation : la somme des produits code \* signal reçu sur la durée d'un bit.
- Les codes sont choisis tels que leur produit scalaire  $C_A \cdot C_B$  soit nul et  $C_A \cdot C_A$  soit maximum (codes orthogonaux = produit scalaire nul). On rappelle qu'un produit scalaire est la somme des produits des composantes correspondantes :  $u_1v_1 + u_2v_2$  pour deux vecteurs  $U$  et  $V$  de composantes  $u_1, u_2$  et  $v_1, v_2$ . Cette notion de produit scalaire n'est pas limitée aux vecteurs dans le plan, mais est général. Le code est ici un vecteur dont les composantes sont les chips.
- Si l'on effectue une moyenne non pas sur les chips d'une période mais à cheval, les moyennes seront globalement plus faibles, donc ceci permet de synchroniser la réception pour certains systèmes, en recherchant la position où l'on obtient (en valeur absolue) un maximum. Les codes sont dans ce cas choisis tels que le produit scalaire d'un code par lui-même décalé soit pratiquement nul.

A partir des spécifications techniques de certains standards [9, 10, 11, 12], nous pouvons résumer dans le tableau 2.1 leurs spécificités et leurs similitudes [7] :

standard	Bande de fréquence (MHz)	Débit (bps)	Technique d'accès	Facteur d'étalement
IS-95	824 – 894 896 - 966	1.2288 M	DS-CDMA	256
BLEUTOOTH	2400 – 2483.5	1 M	FH-CDMA	79
UMTS	1900 – 2025 2110 - 2200	3.83 M	DS-CDMA	4,8,.....,256
CDMA 2000	824 – 894 869 - 894	1.22883 M 3.6864 M	DS-CDMA	4,8,.....,128 4,8,.....,256
WLAN	2400 – 2484	11 M	DS-CDMA	13
ZIGBEE	868 – 868.6 902 – 928 2400 – 2483.5	20 K 40 K 250 K	DS-CDMA	1 10 16

**Tableau 2.1. Caractéristiques de quelques standards de télécommunication.**

Le tableau 2.1 montre aussi que la DS-CDMA est la technique dominante dans presque tous les systèmes de 3G.

### 2.4 Présentation formelle de l'étalement de spectre :

Comme indiqué plus haut, l'étalement de spectre constitue la base des techniques d'accès multiple à répartition par les codes. Il repose sur l'utilisation de séquences pseudo aléatoires, ou codes orthogonaux entre elles au sens des fonctions. Chaque symbole en bande de base de durée  $T_s$  issu d'une source quelconque  $S_j$  est multiplié par la séquence  $C_j$ . L'ensemble des  $C_j$  ( $1 \leq j \leq J$ ) constitue une base orthogonale de dimension  $J$  ; alors,  $J$  source peuvent partager la même bande de fréquence par la superposition des  $J$  signaux séparables résultants. Les symboles d'information en bande de base (impulsions rectangulaires) ont une durée  $T_s$ , suivie d'un intervalle de garde  $\delta$ . Cela permet d'éviter les recouvrements entre symboles successifs par suite des chemins multiples [6].

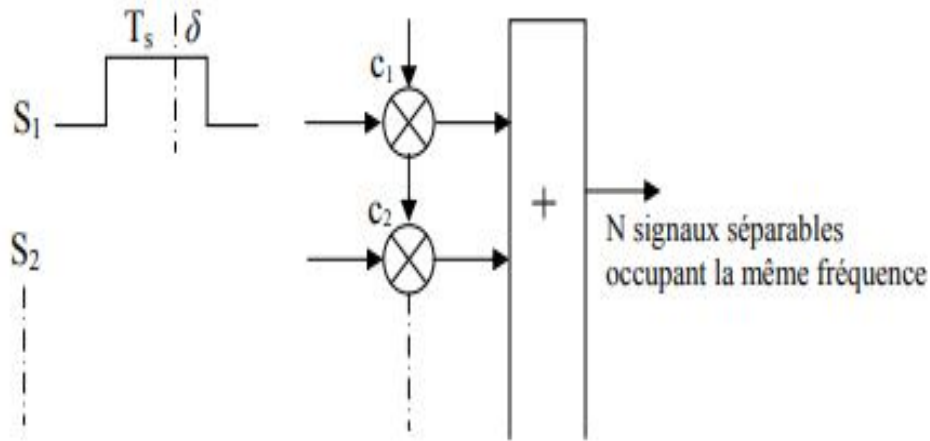


Figure 2.2 : Schéma de principe d'un système de transmission à accès multiple par le code (CDMA).

On dispose d'une suite de symboles informatifs complexes  $\{ \mathbf{b}_k \}$  en bande de base (base bande), provenant d'un usage placé à l'entrée de la liaison numérique. Le symbole courant a pour indice  $k$  et pour durée  $T_s$ . On dispose par ailleurs d'une séquence  $\{ \mathbf{p}_n, n=0 \text{ à } n=L-1 \}$ , dénommée 'code d'étalement', formée de  $L$  éléments appelés 'chips'. Le signal modulé par la séquence correspond à l'émission de la suite :

$$\dots \mathbf{b}_{k-1} \mathbf{p}_0, \mathbf{b}_{k-1} \mathbf{p}_1, \mathbf{b}_{k-1} \mathbf{p}_2, \mathbf{b}_{k-1} \mathbf{p}_{n-1}, \mathbf{b}_k \mathbf{p}_0, \mathbf{b}_k \mathbf{p}_1, \mathbf{b}_k \mathbf{p}_2, \dots, \mathbf{b}_k \mathbf{p}_{n-1}, \mathbf{b}_{k+1} \mathbf{p}_0 \dots$$

Il y a émission d'un nouveau symbole toutes les  $(T_s / L) = T_c$  secondes, au moyen d'une impulsion  $q(t)$  de durée  $T_c$  ( $q(t)=0$  ou  $1$ ). Si la séquence  $\{ \mathbf{p}_n \}$  est suffisamment proche d'une séquence aléatoire, le nouveau signal, modulé, peut être considéré comme un signal de débit  $L$  fois plus grand que le signal initial envoyé par l'utilisateur sur la liaison. Il occupera donc une bande de fréquence  $L$  fois plus étendue, de largeur  $(L / T_s)$  au lieu de  $(1 / T_s)$ . Le débit du signal modulé par le code s'exprime en chips/s.

### 2.4.1 Elaboration analytique du signal temporel d'étalement :

A partir de la séquence de chiffres  $\{ \mathbf{p}_n \}$  et d'une impulsion  $q(t)$  de durée  $T_c = T_s / L$ , on peut construire une séquence temporelle  $c(t)$ , donnée par :

$$c(t) = \sum_{n=0}^{L-1} p_n q(t - nT_c) \tag{Eq. 2.6}$$

Donc la fonction  $q(t - nT_c)$  est la translatée de la fonction  $q(t)$ , dans la translation  $nT_c$ , vers les temps croissants (ou la valeur de  $q(t)$  retardée d'une quantité  $nT_c$ ).

Les schémas qui suivent Figure (2.3) illustrent la relation de définition d'une séquence  $q(t)$  sur un exemple concret de suite  $\{p_n\}$ . Le dernier dessin représente la équation 2.6 devient :

$$c(t) = p_0q(t) + p_1q(t - T_c) + p_2q(t - 2T_c) + \dots$$

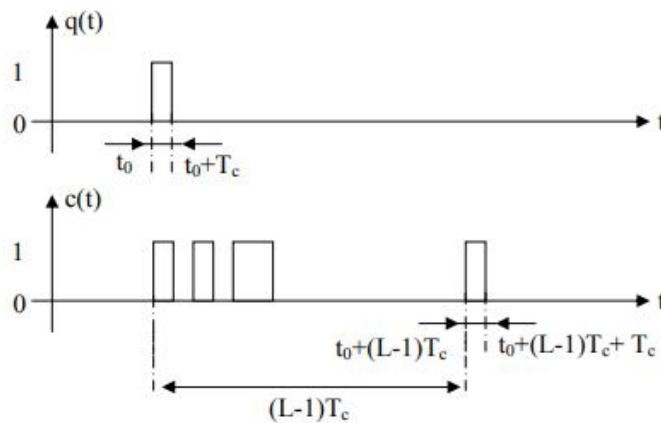
Avec les valeurs explicites suivant :

0	1	(L'indice numérique n )					L-1	n
P0	P1	P2	P3	P4	P5	.....	PL-1	
1	0	1	0	1	1	.....	1	

**Tableau 2.2 : la suit de séquence de chiffré P**

On trouve que  $c(t)$  est :

$$c(t) = 1.q(t) + 0.q(t - T_c) + 1.q(t - 2T_c) + 0.q(t - 3T_c) + 1.q(t - 4T_c) + 1.q(t - 5T_c) + \dots + 1.q(t - (N-1)T_c)$$



**Figure 2.3 : Elaboration d'une séquence d'étalement (code).**

2.4.2 Elaboration analytique du signal temporel étalé :

A la suite  $\{ b_k \}$ , a priori indéfinie, et au signal aléatoire de séquence d'étalement équation 2.6 , on associe un nouveau signal, dit 'étalement' :

$$s(t) = \sum_k b_k c(t - kT_s) \tag{Eq. 2.7}$$

Ce signal est obtenu en multipliant (convoluant) le signal de séquence d'étalement  $c(t)$  sur chacun des intervalles  $T_c$  par le chiffre  $b_k$  correspondant. Si l'on explicite la fonction  $c(t - kT_s)$  , on obtient :

$$c(t - kT_s) = \sum_{n=0}^{L-1} p_n q(t - nT_c - kT_s) \tag{Eq. 2.8}$$

On peut donc préciser l'écriture en appelant  $c_{k,n}(t - kT_s)$  , la séquence qu'il convient de convoluer par le nombre  $b_k$  dans l'intervalle  $kT_s \leq t \leq (k+1)T_s$  . Concrètement, cela veut dire que tous les symboles  $b_k$  (indice  $k$  croissant) vont être convolués par la même période  $T_s$  . Les schémas qui suivent figure (2.4) explicitent ce processus sur un exemple associé à l'exemple précédent [6].

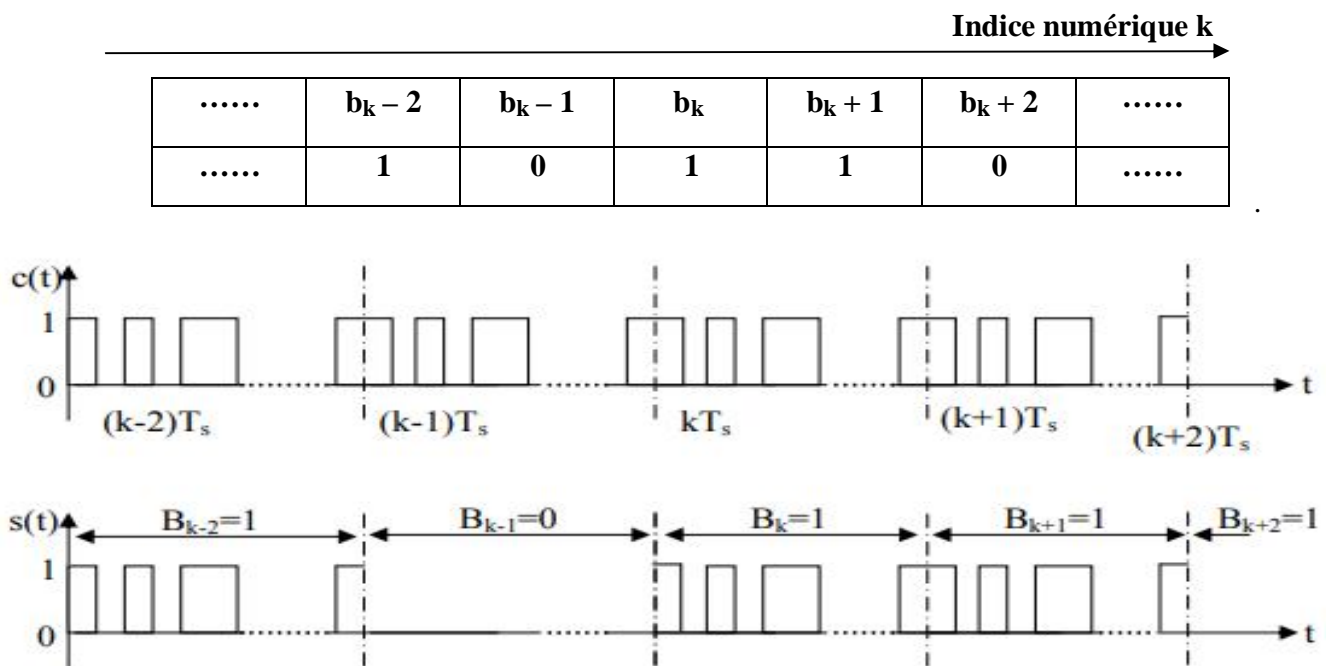


Figure 2.4. Elaboration d'un signal étalé.

Compte tenu des valeurs des deux suites  $\{ p_n \}$  et  $\{ b_k \}$  dans l'exemple étudié, on a bien :

$$S(t) = \dots + 1 \cdot c(t - (k-2)T_s) + 0 \cdot c(t - (k-1)T_s) + 1 \cdot c(t - kT_s) + 1 \cdot c(t - (k+1)T_s) + \dots$$



## 2.5 Les codes d'étalement :

Les communications qui nous intéressent sont les communications radio-mobile. Suivant les caractéristiques de la communication, un choix sur les codes doit être fait.

En communication synchrone, les messages des utilisateurs sont émis dans le canal sans retard relatif. Lorsque le récepteur est synchronisé sur le message reçu, la séparation des utilisateurs peut être faite grâce à ces codes orthogonaux, les codes de Walsh-Hadamard sont dans ce cas les codes les plus utilisés. Lorsque la communication est asynchrone, cas de communication le plus délicat, les messages sont émis avec des retards relatifs qui ne sont pas connus a priori. Dans ce cas, ces codes ne sont plus utilisés, car ils perdent leurs propriétés d'orthogonalité quand ils sont décalés les uns par rapport aux autres. Parmi les codes les plus utilisés pour ce type de communications asynchrones : les codes Gold [13], les codes de JKasaini, spécifiés pour la liaison montante du système UMTS [8], ou les codes Barker [7]. Ces derniers, les codes Barker, sont en nombre très limité et ne sont pas utilisés pour les accès multiples. Les familles de codes cités précédemment sont des codes binaires qui sont utilisés aujourd'hui dans les systèmes CDMA. Il existe cependant d'autres familles de codes non binaires, les codes complexes encore appelés codes polyphasés qui sont, comme les codes binaires, des codes à enveloppes constantes. Parmi ces codes, citons les séquences de Zadoff-Chu ou les séquences de Frank [14]. Ces codes, qui sont nombreux, restent aujourd'hui encore peu appliqués aux systèmes de communications numériques. Notons enfin qu'il est aussi possible de généraliser les séquences binaires en séquences polyphasées [15].

Dans l'interface d'air d'UTRAN (comme exemple des systèmes qui utilise le DS-CDMA) Il ya deux types de codes d'étalement : codes orthogonaux et codes pseudo aléatoires. Des codes pseudo aléatoires sont également connus comme codes de la pseudo-Noise (PN), et les codes orthogonaux utilisés en DS-CDMA sont choisis d'après un ensemble de fonctions proposées par Walsh

### 2.5.1 Codes orthogonaux :

#### 2.5.1.1 Génération de codes Walsh :

Dans la technique d'accès multiple, les utilisateurs partagent le même espace fréquentiel et transmettent sur les mêmes intervalles temporels, alors qu'ils occupent la même bande Radio fréquence. Afin d'éviter les interférences mutuelles 'inter-utilisateurs', les codes Walsh dans IS95 sont utilisés pour séparer les utilisateurs, avec un ensemble de 64 séquences binaires orthogonales [16]. Ces séquences sont orthogonales les unes aux autres, et elles sont générées en utilisant la matrice de Hadamard. La récursivité est utilisée pour générer des matrices d'ordre supérieur de rang inférieur; qui est :

$$H_{2N} = \begin{pmatrix} H_N & H_N \\ H_N & \bar{H}_N \end{pmatrix} \quad (\text{Eq. 2.9})$$

$\overline{HN}$  Contient les mêmes éléments de  $H_N$ , mais inversé. La matrice devient (pour  $N=1$ ):

$$H_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad (\text{Eq. 2.10})$$

Par conséquent, pour tirer une série de quatre séquences orthogonales du code Walsh  $c_0$ ,  $c_1$ ,  $c_2$  et  $c_3$ , nous avons seulement besoin de générer une matrice de Hadamard d'ordre 4, ou :

$$H_4 = \begin{pmatrix} H_2 & H_2 \\ H_2 & \overline{H_2} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Les quatre séquences orthogonales dans cette série de codes de Walsh sont les lignes tirées de la matrice  $H_4$ , donc,

$$\begin{aligned} c_0 &= (0 & 0 & 0 & 0) \\ c_1 &= (0 & 1 & 0 & 1) \\ c_2 &= (0 & 0 & 1 & 1) \\ c_3 &= (0 & 1 & 1 & 0) \end{aligned}$$

Pour DS-SS accès multiple trois conditions doivent être remplies par un ensemble de séquences orthogonales. Ces trois conditions sont :

1. L'inter-corrélation devrait être nulle ou très faible.
2. Chaque séquence dans l'ensemble a un nombre égal de 1s et -1s, ou le nombre de 1s diffère du nombre de -1s.
3. Les codes sont choisis tels que leur produit scalaire doit être égal à 1.

En changeant le 0 à -1 dans chacune des quatre séquences ci-dessus, donnera :

$$\begin{aligned} c_0 &= (-1 & -1 & -1 & -1) \\ c_1 &= (-1 & +1 & -1 & +1) \\ c_2 &= (-1 & -1 & +1 & +1) \\ c_3 &= (-1 & +1 & +1 & -1) \end{aligned}$$

On peut facilement vérifier que toutes les séquences ci-dessus, sauf  $C_0$  remplissent les conditions. . En général, la séquence  $C_0$  du code Walsh se compose uniquement de -1 et ne peut donc pas être utilisées pour la canalisation. Dans le standard IS-95 CDMA,  $C_0$  n'est pas utilisé pour transmettre des informations de base. L'équation (2.9) peut être utilisée récursivement à générer des matrices de Hadamard d'ordres supérieurs afin d'obtenir une plus grande séries de séquences orthogonales. Par exemple, 8 séquences orthogonales, de longueur 8, peuvent être obtenues en générant la matrice  $H_8$ . Dans L'EST-95, le lien descendant utilise une série de 64 séquences orthogonales de Walsh, donc la limitation physique sur le nombre de chaînes sur le lien descendant est de 63 parce que, dans un IS-95,  $C_0$  n'est pas utilisé pour transmettre des informations de base

### Exemple :

L'équation (2.9) peut être utilisée pour générer  $H_8$ , qui est :

$$H_8 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

Les huit codes orthogonaux résultant de Walsh sont :

$$c_0 = (-1 \quad -1 \quad -1 \quad -1 \quad -1 \quad -1 \quad -1 \quad -1)$$

$$c_1 = (-1 \quad +1 \quad -1 \quad +1 \quad -1 \quad +1 \quad -1 \quad +1)$$

$$c_2 = (-1 \quad -1 \quad +1 \quad +1 \quad -1 \quad -1 \quad +1 \quad +1)$$

$$c_3 = (-1 \quad +1 \quad +1 \quad -1 \quad -1 \quad +1 \quad +1 \quad -1)$$

$$c_4 = (-1 \quad -1 \quad -1 \quad -1 \quad +1 \quad +1 \quad +1 \quad +1)$$

$$c_5 = (-1 \quad +1 \quad -1 \quad +1 \quad +1 \quad -1 \quad +1 \quad -1)$$

$$c_6 = (-1 \quad -1 \quad +1 \quad +1 \quad +1 \quad +1 \quad -1 \quad -1)$$

$$c_7 = (-1 \quad +1 \quad +1 \quad -1 \quad +1 \quad -1 \quad -1 \quad +1)$$

### 2.5.1.2 Canalisation en utilisant les codes Walsh :

L'exemple suivant illustre comment les codes Walsh peuvent être utilisés pour DS-SS accès multiple. Supposons qu'il y a trois utilisateurs différents, et chaque utilisateur souhaite envoyer un message distinct. Les messages sont séparés.

$$\mathbf{m1} = (+1 \ -1 \ +1) \quad \mathbf{m2} = (+1 \ +1 \ -1) \quad \mathbf{m3} = (-1 \ +1 \ +1)$$

Chacun des trois utilisateurs est attribué un code Walsh, respectivement :

$$c_1 = (-1 \ +1 \ -1 \ +1)$$

$$c_2 = (-1 \ -1 \ +1 \ +1)$$

$$c_3 = (-1 \ +1 \ +1 \ -1)$$

Chaque message transmis lui a été attribué le code Walsh. Il est à noter que le début de chips du code Walsh est de quatre fois le débit du message, ce qui contribue à un gain de traitement égal à 4. Pour le premier message :

$$m_1(t) = (+1 \quad \quad \quad -1 \quad \quad \quad +1 \quad \quad \quad )$$

$$m_1(t) = (+1 \ +1 \ +1 \ +1 \ -1 \ -1 \ -1 \ -1 \ +1 \ +1 \ +1 \ +1)$$

$$c_1(t) = (-1 \ +1 \ -1 \ +1 \ -1 \ +1 \ -1 \ +1 \ -1 \ +1 \ -1 \ +1)$$

$$m_1(t).c_1(t) = (-1 \ +1 \ -1 \ +1 \ +1 \ -1 \ +1 \ -1 \ -1 \ +1 \ -1 \ +1)$$

Il est à noter que  $m_1(t) \cdot C_1(t)$  qu'est la séquence étalée du signal du premier message. De même, pour le deuxième message:

$$m_2(t) = (+1 \quad \quad \quad +1 \quad \quad \quad -1 \quad \quad \quad )$$

$$m_2(t) = (+1 \ +1 \ +1 \ +1 \ +1 \ +1 \ +1 \ +1 \ -1 \ -1 \ -1 \ -1)$$

$$c_2(t) = (-1 \ -1 \ +1 \ +1 \ -1 \ -1 \ +1 \ +1 \ -1 \ -1 \ +1 \ +1)$$

$$m_2(t).c_2(t) = (-1 \ -1 \ +1 \ +1 \ -1 \ -1 \ +1 \ +1 \ +1 \ +1 \ -1 \ -1)$$

Pour le troisième message:

$$m_3(t) = (-1 \quad \quad \quad +1 \quad \quad \quad +1 \quad \quad \quad )$$

$$m_3(t) = (-1 \quad -1 \quad -1 \quad -1 \quad +1 \quad +1 \quad +1 \quad +1 \quad +1 \quad +1 \quad +1)$$

$$c_3(t) = (-1 \quad +1 \quad +1 \quad -1 \quad -1 \quad +1 \quad +1 \quad -1 \quad -1 \quad +1 \quad +1 \quad -1)$$

$$m_3(t).c_3(t) = (+1 \quad -1 \quad -1 \quad +1 \quad -1 \quad +1 \quad +1 \quad -1 \quad -1 \quad +1 \quad +1 \quad -1)$$

Les étalements du spectre pour les trois messages  $m_1(t).c_1(t)$ ,  $m_2(t).c_2(t)$  et  $m_3(t).c_3(t)$  sont combinés pour former un signal composite  $S(t)$

c'est-à-dire  $S(t)$  est devient :

$$S(t) = m_1(t).c_1(t) + m_2(t).c_2(t) + m_3(t).c_3(t)$$

Le signal  $S(t)$  résultant est:

$$S(t) = (-1 \quad -1 \quad -1 \quad 3 \quad -1 \quad -1 \quad 3 \quad -1 \quad -1 \quad 3 \quad -1 \quad -1)$$

$S(t)$  est le signal composite qui est transmis dans la seule bande RF. S'il ya des erreurs négligeables pendant le processus de transmission, le récepteur intercepte  $S(t)$ . Afin de distinguer les messages originaux  $m_1(t)$ ,  $m_2(t)$  et  $m_3(t)$  à partir du signal composite  $S(t)$ , le récepteur multiplie  $S(t)$  par le code Walsh attribué pour chaque message:

$$S(t).c_1(t) = (+1 \quad -1 \quad +1 \quad 3 \quad +1 \quad -1 \quad -3 \quad -1 \quad +1 \quad 3 \quad +1 \quad -1)$$

$$S(t).c_2(t) = (+1 \quad +1 \quad -1 \quad 3 \quad +1 \quad +1 \quad 3 \quad -1 \quad +1 \quad -3 \quad -1 \quad -1)$$

$$S(t).c_3(t) = (+1 \quad -1 \quad -1 \quad -3 \quad +1 \quad -1 \quad 3 \quad +1 \quad +1 \quad 3 \quad -1 \quad +1)$$

Puis le récepteur intègre, ou ajoute, toutes les valeurs au-dessus de chaque période. Les fonctions  $m_1(t)$ ,  $m_2(t)$  et  $m_3(t)$  sont les résultats:

$$S(t).c_1(t) = (+1 \quad -1 \quad +1 \quad 3 \quad +1 \quad -1 \quad -3 \quad -1 \quad +1 \quad 3 \quad +1 \quad -1)$$

$$M_1(t) = ( \quad \quad \quad 4 \quad \quad \quad -4 \quad \quad \quad 4)$$

$$S(t).c_2(t) = (+1 \quad +1 \quad -1 \quad 3 \quad +1 \quad +1 \quad 3 \quad -1 \quad +1 \quad -3 \quad -1 \quad -1)$$

$$M_2(t) = ( \quad \quad \quad 4 \quad \quad \quad 4 \quad \quad \quad -4)$$

Un "seuil de décision" se penche sur les fonctions intégrées  $m_1(t)$ ,  $m_2(t)$  et  $m_3(t)$ . Les règles de décision sont utilisées :

$$\tilde{m}(t) = 1 \quad \text{if } M(t) > 0$$

$$\tilde{m}(t) = -1 \quad \text{if } M(t) < 0$$

Après l'application de ces règles de décision, nous obtenons les résultats:

$$\tilde{m}_1(t) = (+1 \quad -1 \quad +1)$$

$$\tilde{m}_2(t) = (+1 \quad +1 \quad -1)$$

$$\tilde{m}_3(t) = (-1 \quad +1 \quad +1)$$

### 2.6 Conclusion :

La technique d'accès multiples CDMA basée sur l'étalement de spectre est déjà largement utilisée dans les systèmes de communications radio-mobiles, que ce soit dans les normes américaines (IS 95), japonaises ou, plus récemment, européenne (UMTS). Elle offre une solution plus flexible par rapport à celles des deux techniques TDMA et FDMA.

Dans cette partie nous avons montré le principe de l'étalement de spectre par séquence directe ensuite une présentation formelle de l'étalement de spectre et l'élaboration analytique du signal temporel étalé sont données. Au cours de ce chapitre, nous avons rappelé les différents types de codes, regroupant les séquences de Walsh-Hadamard utilisés dans les systèmes DS-CDMA.

Le canal de propagation peut affecter les performances de tout système de radiocommunication à base de CDMA, car il engendre des dégradations sur le signal transmis. Afin de choisir les techniques de traitement de signal appropriées qui seront mises en œuvre dans la chaîne d'émission et de réception, il faut définir les caractéristiques du canal de propagation. Ces dernières seront présentées dans le chapitre suivant.

# *Chapitre 3:*

*Data mining et Benchmark de  
données*



### **3 . Le Data Mining :**

#### **3.1. Introduction :**

Traduit littéralement par " forage des données ", le Data Mining est un processus est un processus non élémentaire de mises à jour de relations, corrélations, dépendances, associations, modèles, structures, tendances, classes, facteurs obtenus en navigant à travers de grands ensembles de données, généralement consignées dans des bases de données (relationnelles ou pas), navigation réalisée au moyen de méthodes mathématiques, statistiques ou algorithmiques [17]. D'après Le Gartner Group, 1996, ce processus peut être itératif et/ou interactif selon les objectifs à atteindre (Bien que non explicitement contenu dans la définition, on considère le Data Mining comme un processus (le plus automatisé possible) qui va des données élémentaires disponibles dans un Data Warehouse à la décision en apportant à chaque étape de ce processus une plus-value informationnelle qui peut aller jusqu'au déclenchement automatique d'actions en fonction de l'information de synthèse mise à jour. On comprend, derrière le concept du Data Mining l'héritage de l'intelligence artificielle et des systèmes experts. Mais on comprend aussi l'utilisation des méthodes d'analyses des données qui ont pour objet de découvrir des structures, des relations entre faits au moyen de données élémentaires et de techniques mathématiques appropriées. On ne s'étonnera donc pas de trouver au catalogue des méthodes de Data Mining aussi bien les réseaux de neurones, les arbres dits de décision que les méthodes de visualisation multidimensionnelle.

#### **3.2 . Définition du data mining :**

Le terme Data Mining désigne l'analyse de données depuis différentes perspectives et le fait de transformer ces données en informations utiles, en établissant des relations entre les données ou en repérant des patterns. Ces informations peuvent ensuite être utilisées par les entreprises pour augmenter un chiffre d'affaires ou pour réduire des coûts. Elles peuvent également servir à mieux comprendre une clientèle afin d'établir de meilleures stratégies marketing .

Le Data Mining est une composante essentielle des technologies Big Data et des techniques d'analyse de données volumineuses. Il s'agit là de la source des Big Data Analytics, des analyses prédictives et de l'exploitation des données. Découvrez la définition complète du terme Data Mining.

Les logiciels Data Mining font partie des outils analytiques utilisés pour l'analyse de données. Ils permettent aux utilisateurs d'analyser des données sous différents angles, de les catégoriser, et de résumer les relations identifiées. Techniquement, le Data Mining est le procédé permettant de trouver des corrélations ou des patterns entre de nombreuses bases de données relationnelles.

### 3.3 . Processus de data mining :

Il est très important de comprendre que le data mining n'est pas seulement le problème de découverte de modèles dans un ensemble de donnée. Ce n'est qu'une seule étape dans tout un processus suivi par les scientifiques, les ingénieurs ou toute autre personne qui cherche à extraire les connaissances à partir des données. En 1996 un groupe d'analystes définit le data mining comme étant un processus composé de cinq étapes sous le standard CRISP-DM (Cross-Industry Standard Process for Data Mining) comme schématisé ci-dessous (figure 3.1) :



Figure 3.1 : Processus de data mining

#### 3.3.1 .Définition et compréhension du problème :

Dans la plus part des cas, il est indispensable de comprendre la signification des domaines à explorer. Sans cette compréhension, aucun algorithme ne va donner un résultat fiable. En effet, Avec la compréhension du problème, on peut préparer les données nécessaires à l'exploration et interpréter correctement les résultat obtenus [18] .

#### 3.3.2. Collecte des données :

D'après la définition du problème et des objectifs du data mining, on peut avoir une idée sur les données qui doivent être utilisées. Ces données n'ont pas toujours le même format et la même structure. On peut avoir des textes, des bases de données, des pages web, ...etc. Parfois, on est amené à prendre une copie d'un système d'information en cours d'exécution, puis ramasser les données de sources éventuellement hétérogènes (fichiers, bases de données relationnelles, temporelles, ...).

### 3.3.3. Prétraitement :

Les données peuvent contenir plusieurs types d'anomalies : des données peuvent être omises à cause des erreurs de frappe ou à causes des erreurs dues au système lui-même, dans ce cas il faut remplacer ces données ou éliminer complètement leurs enregistrements.

Des données peuvent être incohérentes c.-à-d. qui sortent des intervalles permis, on doit les écarter ou les normaliser. Parfois on est obligé à faire des transformations sur les données pour unifier leur poids. Le prétraitement comporte aussi la réduction des données qui permet de réduire le nombre d'attributs pour accélérer les calculs et représenter les données sous un format optimal pour l'exploration. Dans la majorité des cas, le prétraitement doit préparer des informations globales sur les données pour les étapes qui suivent tel que la tendance centrale des données (moyenne, médiane, mode), le maximum et le minimum, le rang, les quartiles, la variance, ... etc.

Plusieurs techniques de visualisation des données telles que les courbes, les diagrammes, les graphes,... etc., peuvent aider à la sélection et le nettoyage des données. Une fois les données collectées, nettoyées et prétraitées on les appelle entrepôt de données (data warehouse) .

### 3.3.4. Estimation du modèle :

Dans cette étape, on doit choisir la bonne technique pour extraire les connaissances (exploration) des données. Des techniques telles que les réseaux de neurones, les arbres de décision, les réseaux bayésiens, le clustering, ... sont utilisées. Généralement, l'implémentation se base sur plusieurs de ces techniques, puis on choisit le bon résultat. Dans le titre suivant on va détailler les différentes techniques utilisées dans l'exploration des données et l'estimation du modèle.

### 3.3.5. Interprétation du modèle et établissement des conclusions :

Généralement, l'objectif du data mining est d'aider à la prise de décision en fournissant des modèles compréhensibles aux utilisateurs. En effet, les utilisateurs ne demandent pas des pages et des pages de chiffres, mais des interprétations des modèles obtenus. Les expériences montrent que les modèles simples sont plus compréhensibles mais moins précis, alors que ceux complexes sont plus précis mais difficiles à interpréter.

**3.4 . Technique du Data Mining :****3.4.1 .Techniques supervisées :**

Dans la modélisation supervisée, ou prédictive, l'objectif est de prédire un événement ou d'estimer les valeurs d'un attribut numérique continue. Dans ces modèles, il existe des champs où les attributs d'entrée et une zone de sortie ou de la cible. Les champs d'entrée sont également appelés " prédicteurs ", car ils sont utilisés par le modèle pour identifier une fonction de prédiction de champ de sortie.

Les modèles prédictifs sont subdivisés en modèles de classification et d'estimation [20] :

- **Les modèle classification** : Dans ces modèles les groupes ou classes cibles sont connus dès le départ. Le but est de classer les cas dans ces groupes prédéfinis ; en d'autres termes, à prévoir un événement. Le modèle généré peut être utilisé comme un moteur de marquage pour l'affectation de nouveaux cas pour les classes prédéfinies. Il estime aussi un score de propension pour chaque cas. Le score de propension dénote la probabilité d'occurrence du groupe cible ou d'un événement.
- **Les modèle d'estimation** : Ces modèles sont similaires à des modèles de classification, mais avec une différence majeure. Ils sont utilisés pour prédire la valeur d'un champ continu en fonction des valeurs observées des attributs d'entrée.

**3.4.1.1 .Arbre de décision :**

Les arbres de décision fonctionnent en séparant de façon récursive la population initiale. Pour chaque groupe, ils sélectionnent automatiquement l'indicateur le plus significatif, le prédicteur qui donne la meilleure séparation par rapport au champ cible. À travers des cloisons successives, leur objectif est de produire sous-segments purs, avec un comportement homogène en termes de production. Ils sont peut-être la technique la plus populaire de classification. Une partie de leur popularité, c'est parce qu'ils produisent des résultats transparents qui sont facilement interprétables, offrant un aperçu de l'événement à l'étude. Les résultats obtenus peuvent avoir deux formats équivalents. Dans un format de règle, les résultats sont représentés dans un langage simple que les règles ordinaires [21] :

**SI (VALEURS PREDICTIVES) ALORS (RESULTAT CIBLE ET SCORE DE  
CONFIANCE)**

Dans une forme d'arborescence, les règles sont représentés graphiquement sous forme d'arbre dans laquelle la population initiale (nœud racine) est successivement divisé en des nœuds terminaux ou feuilles de sous-segments ayant un comportement similaire en ce qui concerne le champ cible.

Les algorithmes d'arbres de décision constituent selon la vitesse et l'évolutivité. Algorithmes disponibles sont :

- C5.0
- CHAID
- Classification et arbres de régression
- QUEST.

### 3.4.1.2 .Règles de décision :

Ils sont assez semblables à des arbres de décision et de produire une liste de règles qui ont le format des états humains compréhensible:

**SI** (VALEURS PREDICTIVES) **ALORS** (RESULTAT CIBLE ET SCORE DE CONFIANCE).

Leur principale différence par arbres de décision, c'est qu'ils peuvent produire plusieurs règles pour chaque enregistrement. Les arbres de décision génèrent des règles exhaustives et mutuellement exclusives qui couvrent tous les records. Pour chaque enregistrement une seule règle s'applique. Au contraire, les règles de décision peuvent générer un ensemble de règles de chevauchement. Plus d'une règle, avec des prédictions différentes, peut être vraie pour chaque enregistrement. Dans ce cas, les règles sont évaluées, à travers une procédure intégrée, afin de déterminer l'une pour l'évaluation. Habituellement, une procédure de vote est appliquée, qui combine les règles et les moyennes de leurs confidences individuelles pour chaque catégorie de sortie. Enfin, la catégorie ayant la confiance la moyenne la plus élevée est sélectionnée comme la prédiction. Les algorithmes de règles de décision comprennent :

- C5.0
- Liste de décision.

### 3.4.1.3 .Régression :

La régression est la méthode utilisée pour l'estimation des valeurs continues. Son objectif est de trouver le meilleur modèle qui décrit la relation entre une variable continue de sortie et une ou plusieurs variables d'entrée. Il s'agit de trouver une fonction  $f$  qui se rapproche le plus possible d'un scénario donné d'entrées et de sorties

### 3.4.1.4 .Réseaux de neurone :

Les réseaux de neurones sont des puissants algorithmes d'apprentissage automatique qui utilisent des fonctions de cartographie complexe, non linéaire pour l'estimation et classification. Ils sont constitués de neurones organisés en couches. La couche d'entrée contient les prédicteurs ou neurones d'entrée. La couche de sortie comprend dans le champ cible. Ces modèles permettent d'estimer des poids qui relient les prédicteurs (couche d'entrée à la sortie). Modèles avec des topologies plus complexes peuvent également inclure, couches cachées intermédiaires, et les neurones. La procédure de formation est un processus itératif. Enregistrements en entrée, avec des résultats connus, sont présentés sur le réseau et la prédiction du modèle est évaluée par rapport aux résultats observés. Erreurs observées sont utilisés pour ajuster et d'optimiser les estimations du poids initial. Ils sont considérés comme des solutions opaques ou "boîte noir" car ils ne fournissent pas une explication de leurs prédictions. Ils fournissent seulement une analyse de sensibilité, qui résume l'importance prédictive des champs d'entrée. Ils nécessitent une connaissance statistique minimum mais, selon le problème, peut nécessiter un temps de traitement à long pour la formation.

### 3.4.2 .Techniques non supervisées :

Dans les modèles non supervisés ou non orientés, il n'y a pas de champ de sortie, il n'y a que des entrées. La reconnaissance de formes est non orienté, elle n'est pas guidée par un attribut cible spécifique. Le but de ces modèles est de découvrir des motifs de données dans l'ensemble des champs d'entrée

Les modèles non supervisés comprennent :

**Les modèles de dispersion :** Dans ces modèles les groupes ne sont pas connus à l'avance. Au contraire, nous voulons que les algorithmes pour analyser les schémas de données d'entrée et d'identifier les regroupements naturels de données ou de cas. Lorsque de nouveaux cas sont marqués par le modèle de cluster généré ils sont affectés à l'un des groupes révélés.

Les modèles d'association de séquences

- **Les modèles de dispersion :** Dans ces modèles les groupes ne sont pas connus à l'avance. Au contraire, nous voulons que les algorithmes pour analyser les schémas de données d'entrée et d'identifier les regroupements naturels de données ou de cas. Lorsque de nouveaux cas sont marqués par le modèle de cluster généré ils sont affectés à l'un des groupes révélés.
- **Les modèles d'association de séquences :** Ces modèles font également partie de la classe de la modélisation non supervisé. Ils ne comportent pas de prédiction directe d'un seul champ. En fait, tous les champs concernés ont un double rôle, car ils agissent comme des entrées et des sorties en même temps. Des modèles d'association de détecter des associations entre des événements discrets, des produits ou des attributs. Les modèles de séquence détectent des associations au fil du temps.

**3.5. LE BENCHMARKING :****3.5.1. Introduction :**

Le Benchmarking est une riposte imaginée au début des années 80 par des industriels américains pour contrer la supériorité japonaise en s'inspirant des méthodes de leurs adversaires. Xerox en fut l'un des initiateurs.

Le Benchmarking concerne l'ensemble des activités, produits et services liés aux processus d'une organisation privée ou publique, commerciale, de production ou de services. (Développement, conception de produits et/ou services, gestion financière, système d'information, ressources humaines, administration, achats, production, communication, commercial, logistique...)[22] .

**Sa finalité est la mise en œuvre des meilleures pratiques (best practices).**

**3.5.2 .Définition du Benchmarking :**

« Le Benchmarking est une action continue de comparaison d'un processus, produit ou service avec une activité similaire réputée la meilleure, afin de fixer des objectifs et des actions d'amélioration ambitieux mais réalistes, pour devenir et rester le meilleur des meilleurs en un temps raisonnable ».

**Gerald J.BALM 1994 cf. bibliographie**

« Le Benchmarking est la recherche des méthodes les plus performantes pour une activité donnée, permettant de s'assurer une supériorité».

**Robert C.CAMP 1992 cf. bibliographie**

« Le Benchmarking est un processus continu d'évaluation des produits, des services et des méthodes par rapport à ceux des concurrents ou des partenaires les plus sérieux ou des organisations reconnues comme leaders ou chefs de file».

**David Kearns, ex directeur général de Xerox Corporation.**

On dit que : C'est une procédure d'évaluation par rapport à un modèle reconnu, inscrite dans une recherche d'excellence

le Benchmarking est une attitude coopérative qui permet une analyse comparative interne ou externe de plusieurs : concepts, méthodes, outils, processus, produits, services. Il s'appuie sur la collecte et l'analyse d'informations quantitatives et qualitatives et sur la compréhension de la culture de l'organisation partenaire. Il doit faire partie intégrante d'un processus d'évaluations et d'améliorations constantes dont la finalité est de devenir ou de rester le meilleur.»

L'ensemble des contributions, tant des chercheurs et des théoriciens que des praticiens et autres experts, fait ressortir un réel consensus sur l'importance des gains de performances obtenus après la mise en œuvre d'un processus de Benchmarking.

Alors en générale Le Benchmarking est un outil d'amélioration continue de la performance.

**3.5.3 .Le concept de Benchmarking:**

Benchmarking et l'étalonnage des performances compétitives des organisations constituent un instrument de qualité qui a pour objectif l'amélioration continue des processus de management des organisations. Il permet une recherche permanente des meilleures pratiques en se comparant, dans un domaine précis, à d'autres organisations leaders.

Il permet d'accéder aux meilleures pratiques opérationnelles et d'être réactif face à la concurrence exacerbée due à la mondialisation des marchés, à la rapidité des mutations technologiques et à l'évolution des systèmes d'information.

Ce processus d'étalonnage a une double vocation. Il permet d'une part d'analyser la confrontation de sa propre organisation aux pratiques, méthodes, processus et outils des meilleurs, d'autre part il constitue un processus d'apprentissage permanent.

Il a pour objet l'observation et l'analyse des processus décisionnels, organisationnels, industriels et commerciaux. Il fournit des points de référence qui servent de base pour évaluer les performances

L'organisation peut, à partir de ceux-ci, concevoir et/ou améliorer ses processus.

L'organisation doit d'abord procéder à l'analyse et la mesure de ses propres performances. Elle doit « s'étalonner ».

L'étalonnage n'est pas synonyme d'analyse de compétitivité qui, elle, repose sur l'analyse de chiffres, de ratios, d'indicateurs et autres éléments statistiques. Il permet d'identifier les meilleures pratiques et la manière dont elles sont réalisées.

Le Benchmarking révèle le niveau de performance auquel l'organisation peut prétendre. Il lui révèle aussi comment atteindre ce niveau.

**3.5.4. les outils de développements :** dans cette thème de recherche en utilise plusieurs langage de programmation comme : MATLAB, WEKA, JAVA

A /- **MATLAB : (matrix laboratory)** est un langage de script émulé par un environnement de développement, il est utilisé à des fins de calcul numérique. Développé par la société The Math Works, MATLAB permet de manipuler des matrices, d'afficher des courbes et des données, de mettre en œuvre des algorithmes, de créer des interfaces utilisateurs, et peut s'interfacer avec d'autres langages comme le C, C++, Java, et Fortran. Les utilisateurs de MATLAB (environ 4 millions en 2019) sont de milieux très différents comme l'ingénierie, les sciences et l'économie dans un contexte aussi bien industriel que pour la recherche. Matlab peut s'utiliser seul ou bien avec des toolboxes (« boîte à outils »).

Dans notre cas en utilise le MATLAB pour applique la technique de DSCDMA et Extrait son BER a utilisation de relation qui rassemble les différent variable ( nombre d'utilisateurs, longueur de code , Eb , N0 ) ces variables en utiliser pour calculer le BER de DSCDMA .



Tout ces données en utiliser comme des attribue pour le benchmark avec langage de programmation WEKA

**B /- WEKA (31) :** Weka est un logiciel qui implémente plusieurs algorithmes d'apprentissage accessibles via une interface graphique ou d'une librairie Java. Cet outil a été développé par une équipe de chercheurs de l'université de Waikato en Nouvelle-Zélande et est disponible gratuitement (licence GNU) à l'adresse suivante :

**<http://www.cs.waikato.ac.nz/ml/weka/>**

### Objectifs :

- Le weka explorer permet de lancer une méthode à partir d'un fichier de l'extension ARFF
- Les résultats sont mis sous la forme d'un fichier texte normalisé.
- Permet de sélectionner la méthode la mieux adaptée ou la plus efficace.

### Utilisation de l'interface graphique de Weka

Cette simple introduction vous indiquera les étapes à suivre pour utiliser les différents algorithmes disponibles dans le logiciel Weka. Il est fortement recommandé d'explorer plus en profondeur le logiciel par vous-même. Démarrer d'abord le logiciel et sélectionner le mode "Explorer" . C'est le seul mode couvert par cette introduction. La fenêtre de l'explorateur apparaît (voir figure 1) :

Cet onglet permet de charger en mémoire une base de données spécifique. Cliquez sur le bouton "Open file" ① et sélectionner le fichier données **DSCDMA.arff** (arff (Attribute-Relation File Format)).dans le répertoire data. Il s'agit d'une base de données sur les conditions de BER nécessaires pour pratiquer une "bonne" ou "mauvaise" configuration. Une fois chargées en mémoire, les informations sur le fichier sont affichées à l'écran. On voit que la base de données contient 24289 instances de 5 attributs. Plusieurs autres informations sont affichées telles que les Différentes valeurs possibles des attributs et un histogramme représentant la distribution des attributs. Il est Aussi possible d'enlever des attributs en les sélectionnant et en cliquant sur le bouton "Remove"[23].

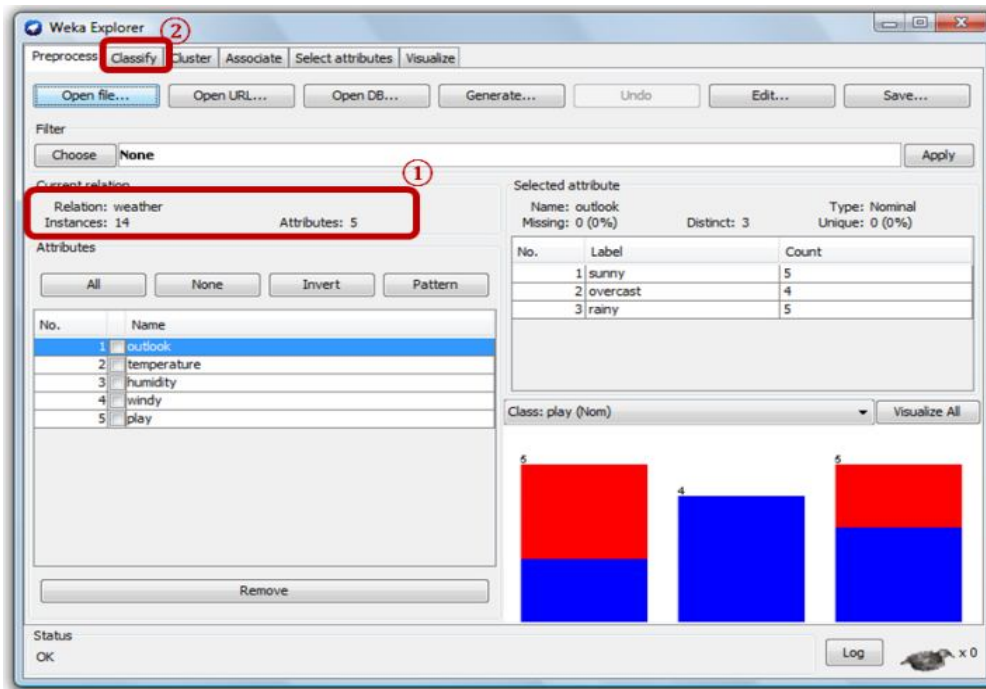


figure 3.2 : Chargement d'une base de données sur Weka

Cliquez sur l'onglet "classify" ② pour la prochaine étape (voir figure 2). L'onglet classificateur affiche une fenêtre qui permet de sélectionner un algorithme d'apprentissage, entraîner le système et le tester à l'aide d'un sous-ensemble de test. Les résultats sont ensuite affichés à l'écran. La première étape consiste à choisir un algorithme d'apprentissage. Appuyez sur le bouton "Choose" ③ et choisissez l'algorithme J48 dans la section "Trees". Cet algorithme est une implémentation légèrement améliorée de l'algorithme C4.5 pour la construction d'arbres de décision. En cliquant sur la barre des propriétés ④ de l'algorithme avec le bouton droit de la souris, une fenêtre avec différentes options pour l'algorithme apparaît. Laissez les options par défaut. Une deuxième section de la fenêtre ⑤ permet de sélectionner les options utilisées lors de la phase de test.

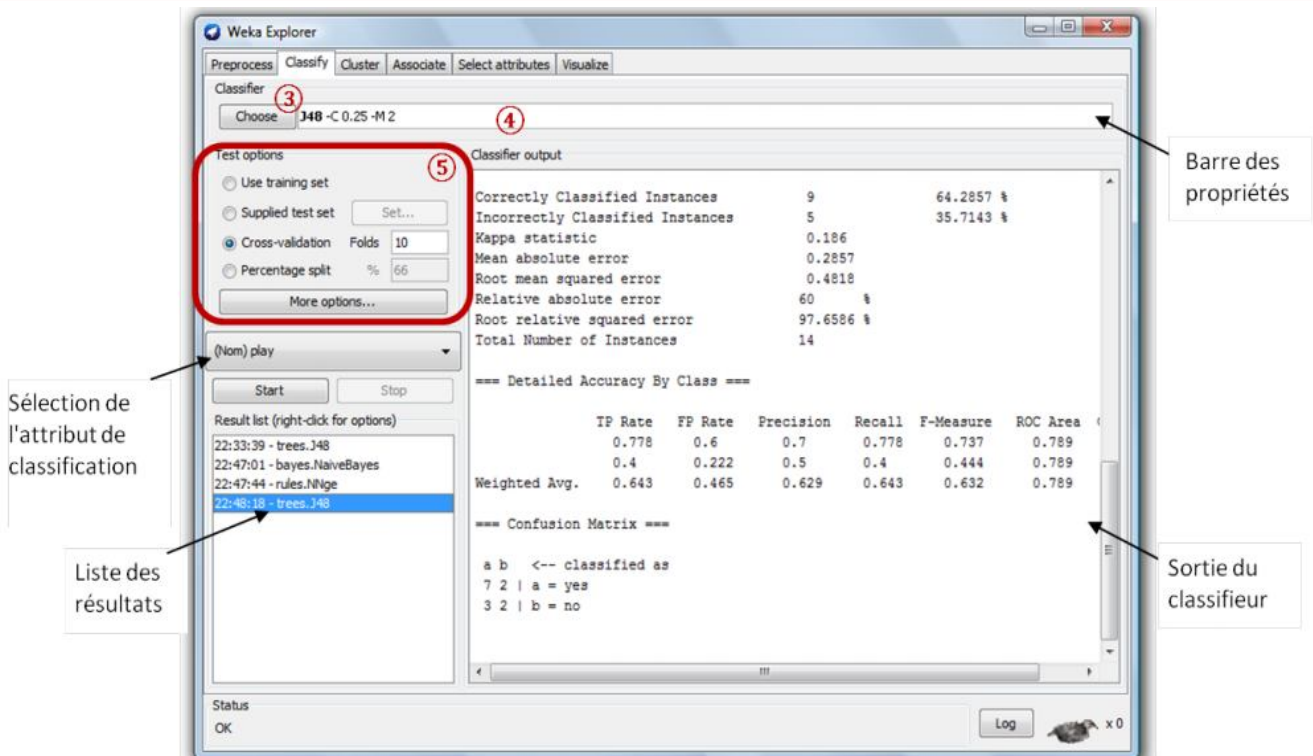


Figure 3.3 : Phase d'apprentissage et test.

**C /- JAVA :** Java est un langage de programmation et une plate-forme informatique qui ont été créés par Sun Microsystems en 1995. Beaucoup d'applications et de sites Web ne fonctionnent pas si Java n'est pas installé et leur nombre ne cesse de croître chaque jour. Java est rapide, sécurisé et fiable. Des ordinateurs portables aux centres de données, des consoles de jeux aux superordinateurs scientifiques, des téléphones portables à Internet, la technologie Java est présente sur tous les fronts .

**3.5.5 .Résultat de simulation :**

**3.5.5.1 .A .Collecte et préparation des données :**

On a recueilli l'ensemble des données dans l'entrepôt de «Machine Learning Repository » qui fournit un ensemble de données destiné à l'apprentissage artificiel et à la fouille de données. Le tableau ci-après montre les caractéristiques des bases choisies :

	Nombre d'instances	Nombre d'attribue	Type d'attribue	Nombre de classes	Description
DSCDMA.arff	<b>24289</b>	<b>5</b>	Réel	<b>2 (bonne, mauvaise)</b>	On cherche dans ces données en fonction de leur attribue pour les quel il est bonne ou mauvaise configuration

**Tableau 3.1 : caractéristiques des tableaux de données DSCDMA.arff**

Dans ce fichier (DSCDMA.arff) il existe 24289 instances avec 5 attribue (**nombre d'utilisateur, longueur de code, énergies de bits, puissance de bruit, BER**) tous ces attribue en utilise pour calculer le BER qui utiliser pour déterminer la classe de résultat elle est **bonne** ou **mauvaise** configuration.

**B /- Représentation des données :**

Les données sont stockées dans des fichiers textes au format **arff** (Attribute-Relation File Format). On a des classes et d'un ensemble d'attributs. Ci-après un extrait d'une des deux bases utilisées qui contient partie l'en-tête et partie données (Data) .

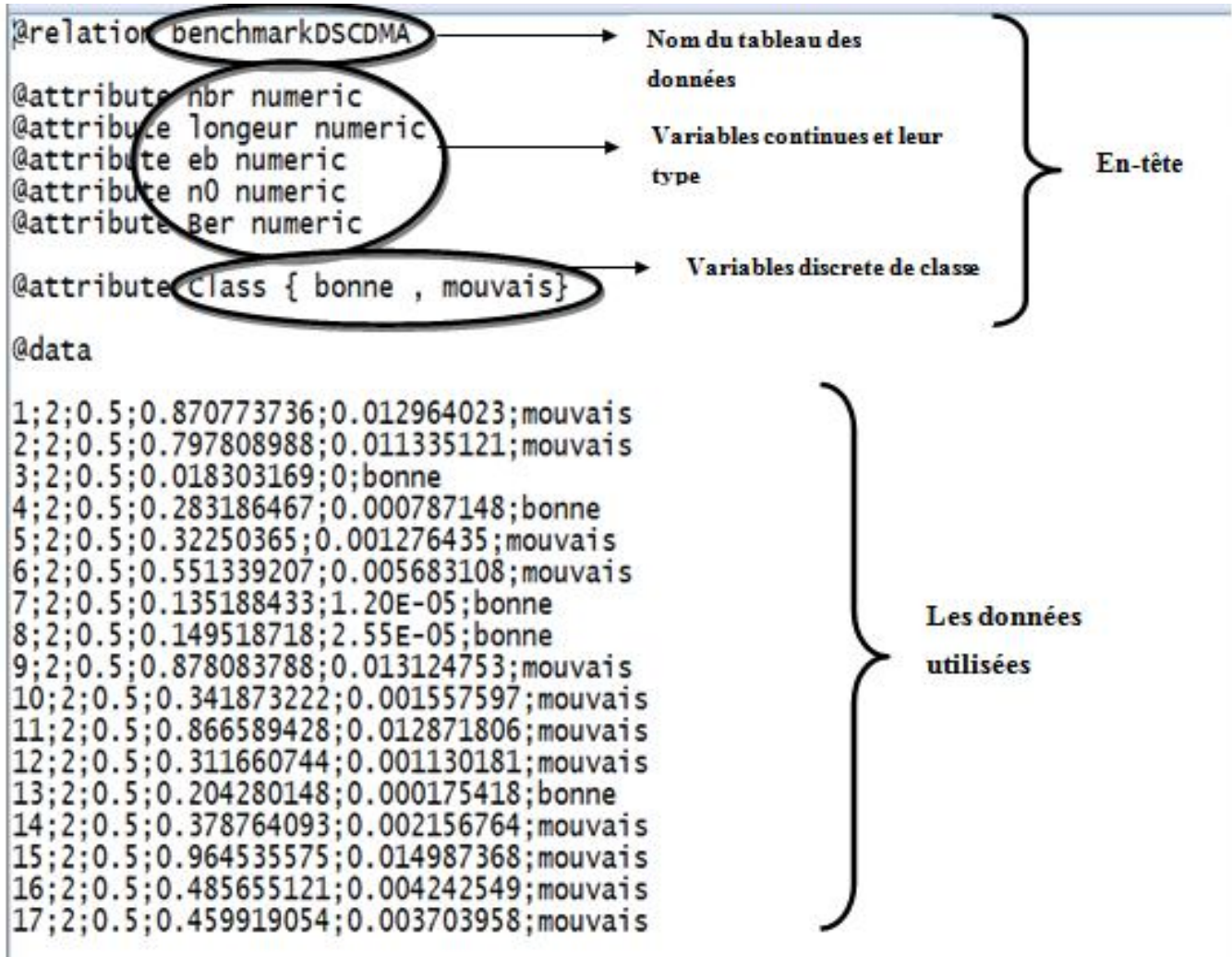


Figure 3.4 : préparation les données de benchmark pour le Weka

3.5.5.2 .Les ensemble de donnés utiliser dans notre benchmark :

Les attribues utilise dans notre benchmark est :

- **Le nombre d'utilisateur U** : le nombre d'utilisateur est 1000000 cette attribue varié entre [ 1 , 2 , 3 , ..., 1000000 ]

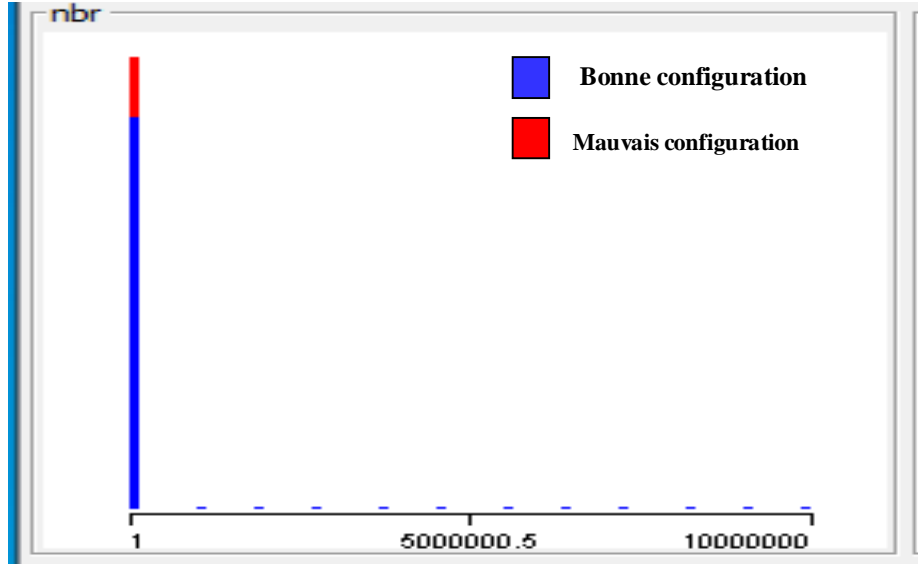


Figure 3.5 : représentation de nombre d'utilisateur sur Weka

- **La longueur de code** : la Longueur de code est varié entre [ 2 , 4 , 8 , .. , 1024 ]

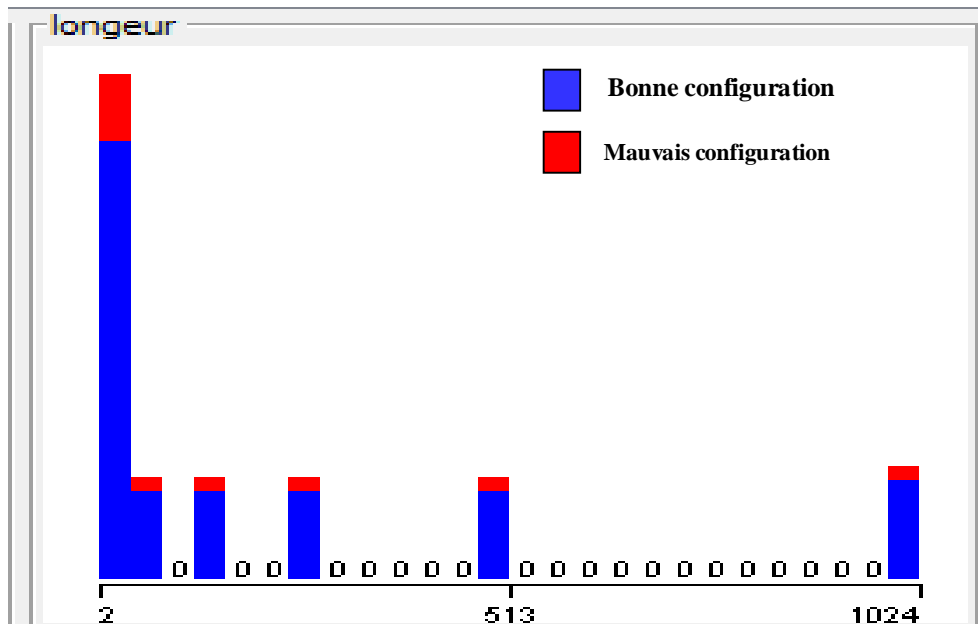


Figure 3.6 : représentation longueur de code sur Weka

3.5.5.3 .Commentaire sue le figure 3.5 et 3.6 :

A /- Premièrement, en ce qui concerne le nombre d'utilisateurs et longueur de code , nous notons qu'il y a eu une diminution du nombre d'utilisateurs , il y a eu une bonne transmission des informations avec peu de présence de mauvaise transmission d'informations, et plus le nombre d'utilisateurs, moins la transmission d'informations et le pourcentage d'erreur qui provoque une mauvaise transmission augmente .

En suit en calcule Le rapport signal sur bruit ( $E_b/N_0$ ) est un indicateur de la qualité de la transmission d'une information qui déterminer le BER . C'est le rapport des puissances entre :

- **L'énergie de bit  $E_b$**  : c'est signal d'amplitude maximale pour laquelle la distorsion à la sortie reste inférieure à une valeur limite son intervalle entre [ 0.5 jusqu'à 2 ] .

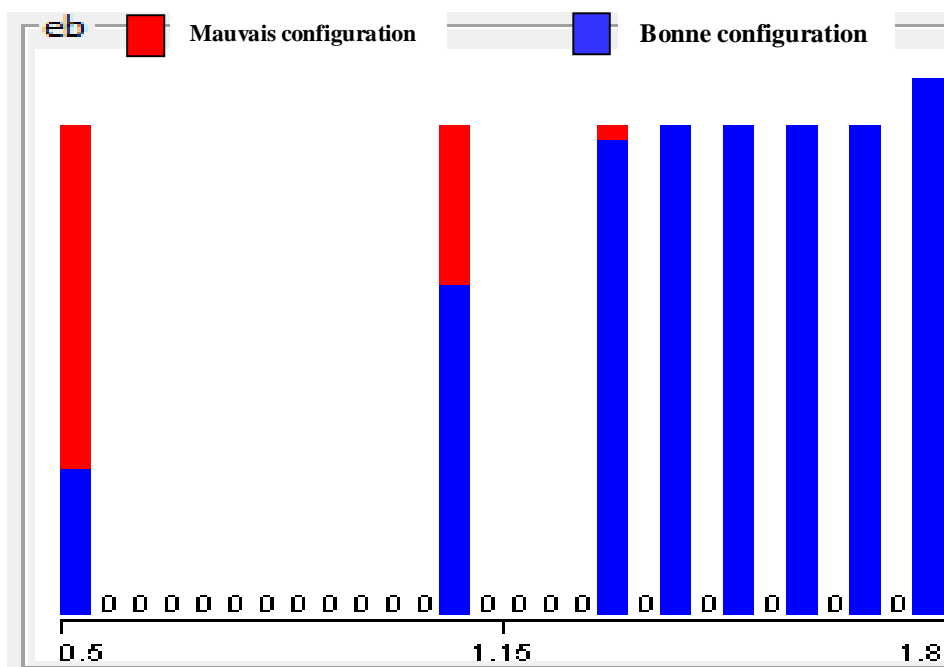


Figure 3.7 : représentation Energie de bit sur Weka

B /- **Commentaire de Figure 3.7** : concernant L'énergie de bit on note que plus il y a plus d'énergie, il y a une bonne transmission de l'information, et c'est clair où entre 0.5 et 1.15 on remarque une transmission de l'information avec quelques mauvaises transmissions. Quant à de 1.15 avec des valeurs plus élevées à 1.8 on constate une très bonne transmission de l'information là où la mauvaise transmission est quasi inexistante .

- **rapport de densité spectrale de puissance de bruit ou (le bruit de fond ) N0** : information non significative correspondant en général au signal présent à la sortie du dispositif en l'absence d'un signal à l'entrée. Pour la valeur de N0 varié entre [ 0 et 1 ] .

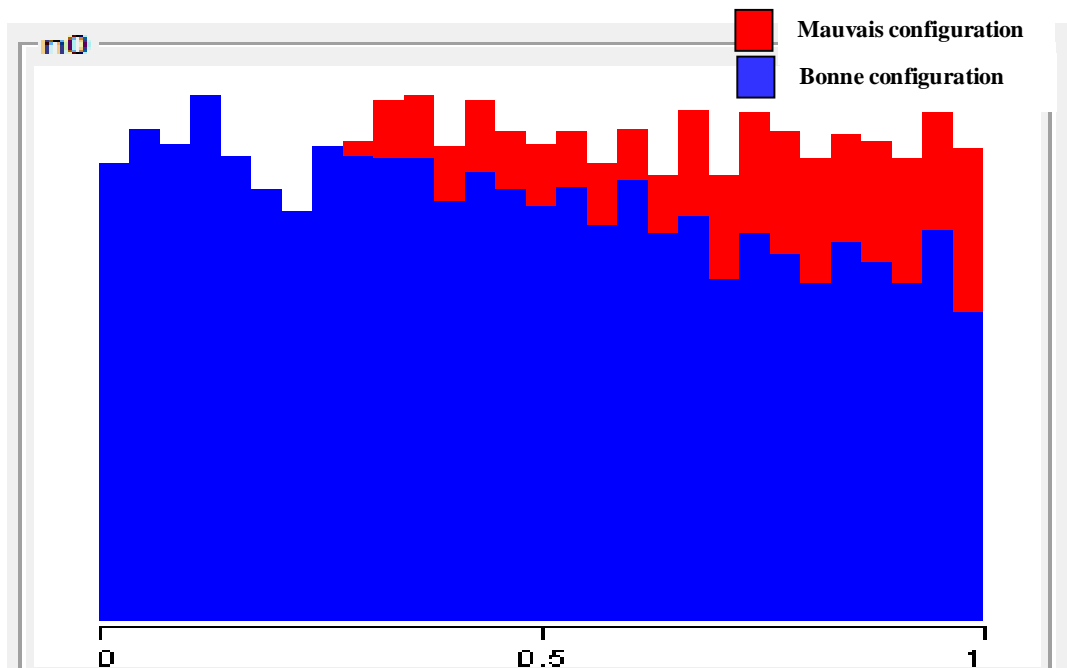


Figure 3.8 : représentation de puissance de bruit sur Weka

**C /- commentaire de Figure 3.8 :** Quant à remarque puissance de bruit N0 , on note que plus elle diminue, alors il est sûr que ce sera une bonne transmission car elle est liée à bruit Et contraire c'est vrai Ceci est entièrement illustré dans un figure où l'on remarque l'apparition d'une mauvaise transmission d'environ 0,4 et continue d'augmenter plus la valeur est élevée.



- BER** : Le taux d'erreur ou B.E.R (**Bit Error Rate** ), désigne la valeur mesurée à la réception d'une transmission numérique, relative au niveau d'atténuation et/ou de perturbation d'un signal transmis , souvent exprimé en pourcentage la probabilité d'erreur de bit  $P_e$  , est une mesure de performance sans unité .

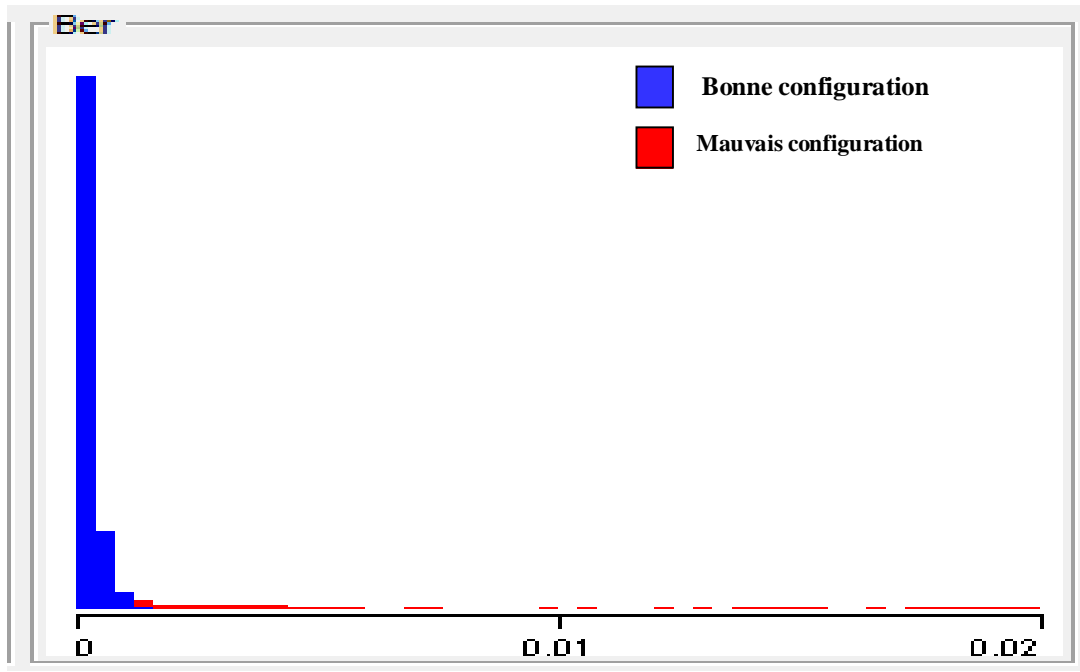


Figure 3.9 : représentation BER (Bits Erreur Rate ) sur Weka

**D/- Commentaire de Figure 3.9 :** Comme le montre cette figure que les résultats ont été impressionnants et c'est une vision de BER , Qui l'a expliqué quand c'est le cas BER Sa valeur est petite, la transmission est très bonne, et plus sa valeur est élevée, plus la transmission devient mauvaise et ça Quesque en veut montre dais tout ca fait d'une manière intelligence .

Les classes de données : si bonne ou mauvaise configuration

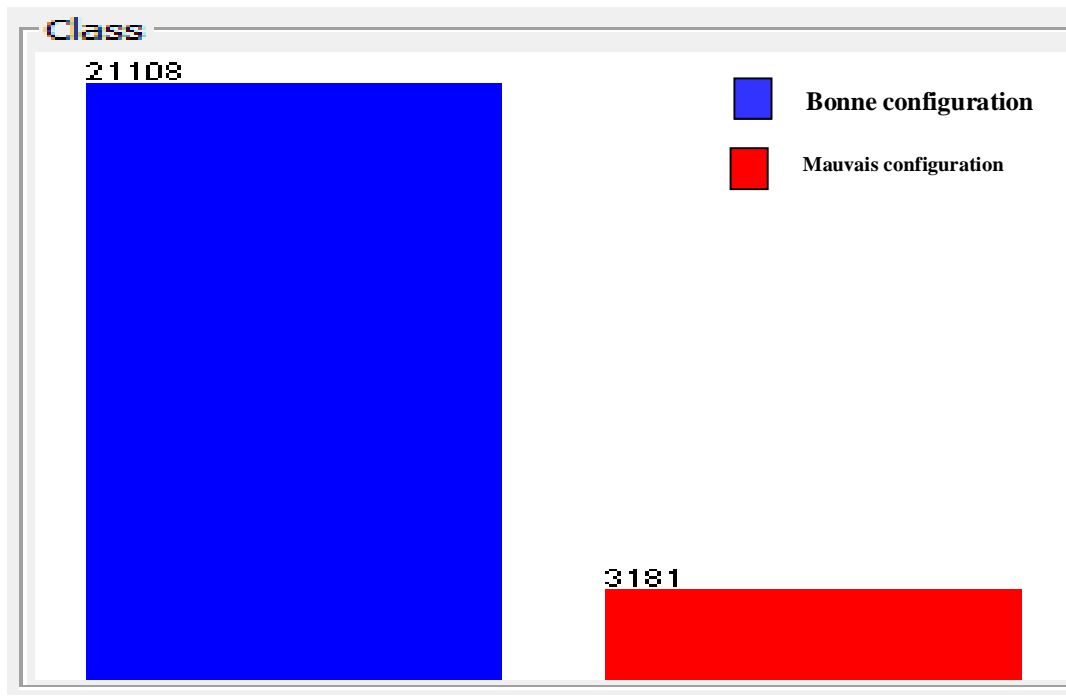


Figure 3.10 : représentation de classe (bonne / mauvaise) sur Weka

**E /- Commentaire de Figure 3.10 :** Toutes ces données précédentes nous ont donné des résultats très impressionnants en ce qui concerne une bonne communication, car elles nous ont donné 21108 comme bonne configuration sur 24289 et le reste, ce qui représente 3181 comme une mauvaise configuration.

**3.5.6 .Interprétation des résultats :**

**On a**  $21108 / 24289 = 0.86904$  qui représente **86.904 %** pour les bonnes configurations

$3181 / 24289 = 0.13096$  qui représente **13.096 %** pour les mauvaises configurations

Dans notre cas précis, on voit que nous obtenons une précision de **86.904 %**. Ce résultat très élevé qu'est mieux lorsqu'on compare avec le hasard. Il est aussi possible de voir le graphique de l'arbre de décision générée par l'expérience en cliquant sur l'expérience dans la liste des résultats avec le bouton de droite et en sélectionnant " Visualize tree ". Différentes options sont disponibles en fonction des classificateurs utilisés. La fenêtre "Classifier Output" contient également des résultats d'autres critères qui sont aussi utilisés pour évaluer les performances du système, nous intéressent par l'Arbre de Décision .

**3.5.7 .L'Arbre de Décision :**

L'apprentissage par arbre de décision désigne une méthode basée sur l'utilisation d'un arbre de décision comme modèle prédictif. On l'utilise notamment en fouille de données et en apprentissage automatique.

Dans ces structures d'arbre, les feuilles représentent les valeurs de la variable-cible et les embranchements correspondent à des combinaisons de variables d'entrée qui mènent à ces valeurs. En analyse de décision, un arbre de décision peut être utilisé pour représenter de manière explicite les décisions réalisées et les processus qui les amènent. En apprentissage et en fouille de données, un arbre de décision décrit les données mais pas les décisions elles-mêmes, l'arbre serait utilisé comme point de départ au processus de décision.

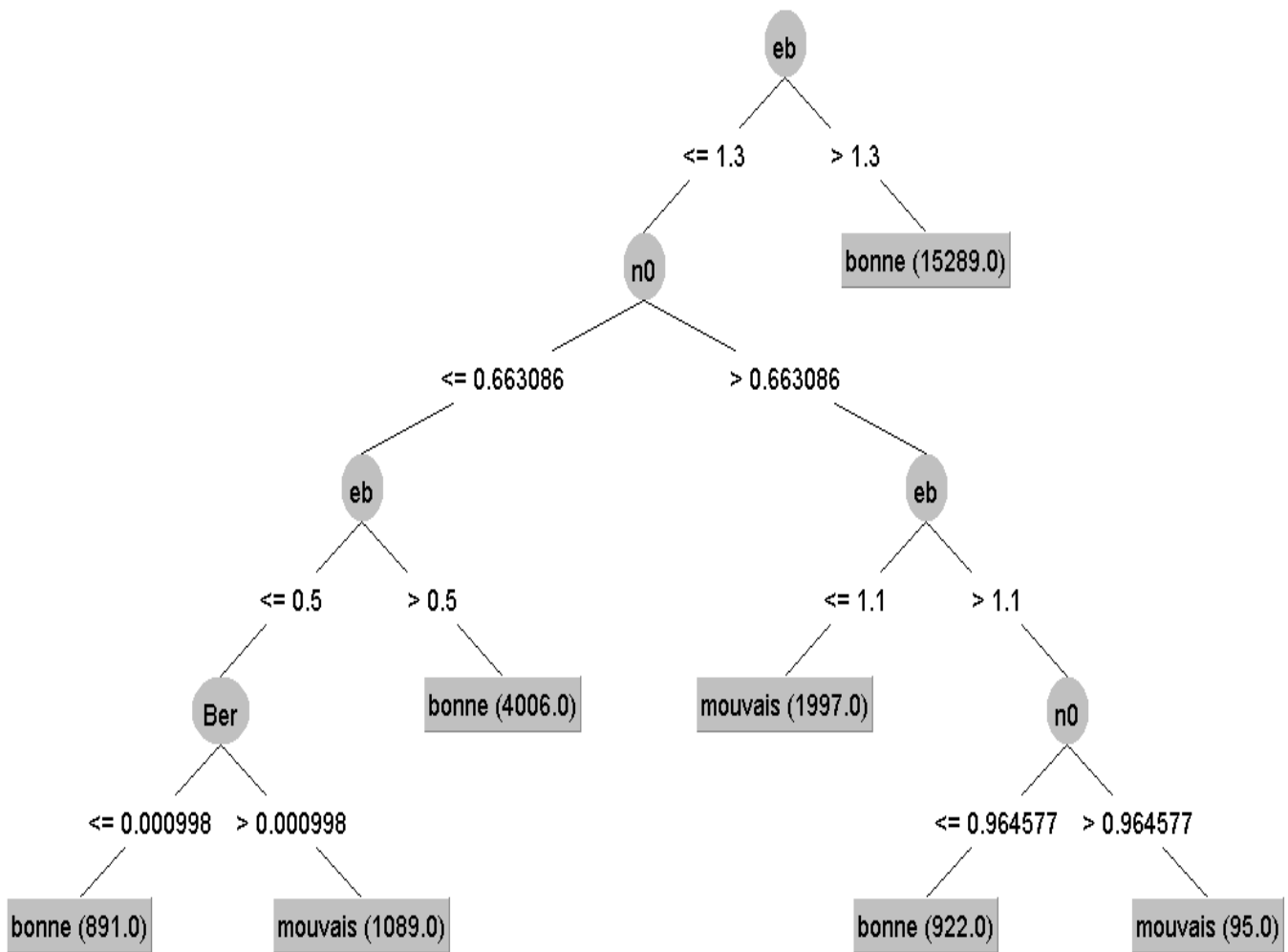


Figure 3.11 : arbre de décision pour notre technique d'accès DSCDMA

### 3.5.8 .Analyse de l'arbre:

D'après la Figure 3.11 qui correspond aux classes et attributs, on constate que tous les attributs choisis de l'ensemble de données étudié apparaissent dans l'arbre de décision. Cela confirme l'importance de ces attributs pour prendre une décision dans le processus de détection des anomalies. En outre, le (**Energie de bit** ) **Eb** défini par l'arbre, du nœud racine signifie qu'il est plus fiable par rapport a les autre attribue . Chaque chemin de l'arbre qui va de la racine à une feuille peut être traduit en règle sous forme :

**Si** conditions **Alors** Décision.

À partir de la Figure 3.11, on peut extraire les règles de décision suivantes :

Règle 1 : Si  $E_b > 1.3$  Alors la configuration est bonne directement sont passé par les autre attribue cette action Contient 15289 instances

On deuxième racine de l'arbre pour que  $E_b \leq 1.3$  on doit voir les valeurs de  $N_0$  par rapport a la valeur 0.663086 pour que Règle 2 :

Règle 2 : Si  $E_b \leq 1.3$  ET  $N_0 > 0.663086$  ET on voir valeur de  $E_b \leq 1.1$  Alors La configuration est mauvaise cette action Contient 1997 instances

De la même manière, on peut traduire l'ensemble des chemins sous forme de règles de décision.

Enfinement nous pouvons être utilisés cet arbre dans les suivis configuration de différent instances pour connaître les meilleurs paramètres des attribue pour que configuration doit être bonne et plus performant possible, ses règles sont très riche connaissance qui peut être facilité les taches aux experts dans le domaine.

**3.6 .Conclusion :**

Le logiciel Weka est un logiciel très puissant pour les besoins d'étude ou de recherche dans le domaine de la prévision et de l'apprentissage automatique des données. Il possède une pluralité de d'algorithmes de filtrage de classification et d'apprentissage qui varient les uns des autres selon leur utilisation et leur performance.

Après ces expériences nous avons obtenus des très bonnes résultats qui représente 86.904 % pour les bonnes configurations cette résultat grâce a l'imputation des attribue (données) à l'aide de l'arbre de décision.

## Conclusion Générale :

Les communications sans fil connaissent une croissance vertigineuse pratiquement partout dans le monde ces dernières années. Pour cela, les études sont devenues axées sur l'amélioration de la qualité des communications radio afin d'assurer une meilleure qualité de transmission, un haut débit et la confidentialité des utilisateurs.

L'apparence de la technique DS-SS-SS-SS a offert plusieurs avantages aux communications radio mobile.

Dans notre travail, nous avons conclu que nous à intégrer deux domaines de travail différents pour développer le système DSSSS, qui sont au cœur de notre travail: le premier c'est le domaine de Télécommunication Ce qui nous importe, c'est la technique de DSSSS, Le deuxième domaine est Data Mining Utilisation d'un programme Weka. Les résultats de notre travail ont été très excellents, car nous avons obtenu de l'aide Data Mining Au pourcentage de 86.904 % d'une bonne configuration , et ce résultat nous pouvons en déduire qu'il y a une grande relation entre les domaines de Télécommunication et Data Mining , qui était notre objectif depuis le début, qui était de développer , renforcer et enrichir notre système .DSSSS avec un grand pourcentage de bonne configuration et pour cela en concluant qu'on a d'atteindre les objectifs fixés .

## Référence

- [1] [http://theses.univ-batna.dz/index.php/theses-en-ligne/search\\_result](http://theses.univ-batna.dz/index.php/theses-en-ligne/search_result)
- [2] <http://di.univ-blida.dz:8080/jspui/bitstream/123456789/2127/1/format-m%C3%A9moire-PFE%2012fi.pdf>
- [3] <http://www.univ-bejaia.dz/dspace/bitstream/handle/123456789/13606/OUSSID%20BOBO.pdf?sequence=1&isAllowed=y>
- [4] <http://dspace.univ-msila.dz:8080/xmlui/bitstream/handle/123456789/105-6789/10594/M2012.14.pdf?sequence=1&isAllowed=y>
- [5] [http://theses.univ-batna.dz/index.php/theses-en-ligne/doc\\_download/2016-analyse-des-methodes-degalisation-des-techniques-cdma-](http://theses.univ-batna.dz/index.php/theses-en-ligne/doc_download/2016-analyse-des-methodes-degalisation-des-techniques-cdma-)
- [6] René Parfait, les réseaux de télécommunication, Lavoisier, 2002
- [7] Amel Aissaoui, ‘Synchronisation Adaptative du code PN Dans les système de communication DS/SS ’, juin 2016.
- [8] <https://pastel.archives-ouvertes.fr/pastel-00001398/document>
- [9] 3GPP-201, ‘TS.25.201 UMTS; Physical layer-general description’ 3GPP Technical Specification, Tech. Rep. March 2001, version 4.0.0.
- [10] IEEE 11.B, ‘IEEE Standard 802.11b,’ IEEE Standardization, Tech. Rep., 1999.
- [11] IEEE 15.1, ‘IEEE Standard 802.15.1, Specification of the Bluetooth system, version 1.2,’ IEEE Standardization, Tech. Rep., November 2003



# Bibliographie

---

- [12] IEEE 15.4, 'IEEE Standard 802.15.4,' IEEE Standardization, Tech. Rep., May 2003.
- [13] R. Gold, 'Optimal binary sequences for spread multiplexing,' IEEE Tran. on information theory, vol. 13, pp. 619-621, Bibliographies 132
- [14] B. M Popovic, 'Generalized chirp-like polyphase sequences with optimum correlation properties,' IEEE Train. On information theory, vol. 38, pp. 1406-1409.
- [15] S. C lui and J. J. Kome, 'Non binary Kazami sequences over GF(p), ' IEEE Train. on information theory, vol. 38, pp. 1409-1410.
- [16] Samuel C. Yang. 'CDMA RF System Engineering' Artech House **Boston**, London.
- [17] STÉPHANE, Tufféry. Data mining et statistique décisionnelle: l'intelligence des données. Editions Technip, 2012.
- [18] <http://depot-e.uqtr.ca/id/eprint/1423/1/030000495.pdf>
- [19] LAJNEF, Mohammed Ali, AYED, Mounir Ben, et KOLSKI, Christophe. Convergence possible des processus du data mining et de conception-évaluation d'IHM: adaptation du modèle en U. In : Proceedings of the 17th Conference on l'Interaction Homme-Machine. 2005. p. 243-246.
- [20] LEE, Yuchun et KENNEDY, Ruby. Visual presentation technique for data mining software. U.S. Patent No 6,269,325, 31 juill. 2001.
- [21] TUFFÉRY, Stéphane. Data mining et statistique décisionnelle: l'intelligence dans les bases de données. Editions Technip, 2005.
- [22] [https://www.performance-publique.budget.gouv.fr/sites/performance\\_publicque/files/files/documents/performance/controle\\_gestion/analyse\\_comparative/concepts\\_et\\_methodologie/Guide\\_methodologique\\_benchmarking\\_Nevaconseil\\_2005.pdf](https://www.performance-publique.budget.gouv.fr/sites/performance_publicque/files/files/documents/performance/controle_gestion/analyse_comparative/concepts_et_methodologie/Guide_methodologique_benchmarking_Nevaconseil_2005.pdf)
- [23] <https://cours.etsmtl.ca/gti770/private/labs/Lab1/LOG770-Labo1-Enonce.pdf>