Université de Montréal

Vie privée en commerce électronique

par
Flavien Serge Mani Onana

Département informatique et de recherche opérationnelle
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en informatique

Décembre 2005

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

Vie privée en commerce électronique

présentée par :

Flavien Serge Mani Onana

a été évaluée par un jury composé des personnes suivantes :

Alain Tapp
président-rapporteur

Esma Aïmeur
directeur de recherche

Gilles Brassard
codirecteur

Douglas Eck
membre du jury

Bart Preneel
examinateur externe

Jian-Yun Nie
représentant du doyen de la FES

Thèse acceptée le 19 mai 2006

RÉSUMÉ

En commerce électronique, les vendeurs se servent de technologies de plus en plus invasives (formulaires de saisie, *spywares*, *cookies*, *Web bugs*, *log files*, etc.) pour colliger des données privées sur leurs clients. Ils justifient généralement ce comportement par deux raisons : d'une part, ils font valoir la nécessité d'offrir de meilleurs services à leurs clients et, d'autre part, ils disent vouloir se mettre à l'abri des éventuelles actions malveillantes des usagers qui se connectent à leurs sites Web commerciaux. Malheureusement, ces éléments de justification ne constituent souvent que le côté visible de l'iceberg. En effet, un vendeur sans scrupules pourrait par exemple partager, échanger, voire commercialiser les données privées colligées sur des clients. Il s'agit là d'une situation qui pourrait conduire à la violation de leur vie privée.

Nous croyons que la vie privée est un droit fondamental pour tous les individus. Dans cette thèse, nous présentons une approche visant la lutte contre la violation de la vie privée des clients, ainsi que la protection des données sensibles des vendeurs en commerce électronique. Pour cela, nous avons procédé, d'une part, à la séparation des données appartenant à chacune des deux entités (clients et vendeurs) et, d'autre part, nous avons créé un ensemble de protocoles et mécanismes destinés à tout processus d'achat/vente en commerce électronique. Ces protocoles et mécanismes sont essentiellement basés sur des techniques cryptographiques.

Mots clés: Vie privée, Cryptographie, Commerce électronique, Masquage, Tiers de semi-confiance, Centre de livraison anonyme.

ABSTRACT

In the context of electronic commerce (e-commerce), merchants use increasingly invasive technologies such as forms, spywares, cookies, Web bugs, log files, etc., to gather information about customers. They put forward two reasons to justify their attitude: first, they claim the necessity for them to offer the best possible services to their customers and second, they want to be protected against potential abuses from users who connect to their commercial Web sites. Unfortunately, this justification is often only the tip of the iceberg. Indeed, an unscrupulous merchant could, for example, share, exchange or even sell information about customers with other merchants and/or organizations, and this may lead to the violation of customer privacy.

We consider that privacy is a fundamental right for every individual. In this thesis, we present an approach to fight against the violation of customer privacy, as well as to protect the merchant's sensitive data in e-commerce. For this purpose, we separated the data belonging to each entity (customers and merchants) and we created protocols and introduced mechanisms intended for any buying/selling process in electronic commerce. These protocols and mechanisms are essentially based on cryptographic techniques.

Keywords: Privacy, Cryptography, Electronic commerce, Blinding, Semi-trusted third party, Anonymous delivery centre.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES APPENDICES

LISTE DES SIGLES

ADC	Anonymous Delivery Centre
AES	Advanced Encryption Standard
ASP	Active Server Pages
BCBB	Blind Customer Buying Behaviour
BliM	Blind Maintenance
BliN	Blind Negotiation
BliP	Blind Payment and delivery
BliS	Blind Search
CA	Certification Authority
CBB	Customer Buying Behaviour
CF	Collaborative Filtering
CGI	Common Gateway Interface
CN	Content-Based Filtering
CNIL	Commission Nationale de l'Informatique et des Libertés
CRL	Certificate Revocation List
CRM	Customer Relation Management
DDHP	Decision Diffie-Hellman Problem
DDoS	Distributed Denial-of-Service
DES	Data Encryption Standard
DF	Demographic Filtering
DLP	Discrete Logarithm Problem
DSA	Digital Signature Algorithm
DSS	Digital Signature Standard

EDI	Electronic Data Interchange
EFT	Electronic Fund Transfers
EPIC	Electronic Privacy Information Center
FAI	Fournisseur d'Accès Internet
HTTP	HyperText Transfer Protocol
IDEA	International Data Encryption Algorithm
IDS	Intrusion Detection System
IETF	Internet Engineering Task Force
IPsec	IP security protocol
ISO	International Organization for Standardization
ISP	Internet Service Provider
KDC	Key Distribution Centre
kNN	k-Nearest Neighbours-based algorithm
KF	Knowledge-Based Filtering
LAN	Local Area Network
LPRP	Loi sur la Protection des Renseignements Personnels
LPRPDÉ	Loi sur la Protection des Renseignements Personnels et les Documents Électroniques
MAC	Manipulation Authentication Codes
MDC	Manipulation Detection Codes
MIC	Message Integrity Codes
MPC	Multi-Party Computation
OCDÉ	Organisation de Coopération et de Développement Économiques

OSI	Open Systems Interconnection
OT	Oblivious Transfer
PET	Privacy Enhancing Technologies
PGP	Pretty Good Privacy
PIR	Private Information Retrieval
PKI	Public Key Infrastructure
PRÉ	Procédure de Remplacement d'État
PRIME	Privacy and Identity Management for Europe
PSTN	Public Switched Telephone Network
P3P	The Platform for Privacy Preferences Project
SLP	Stop List for Privacy
SMTP	Simple Mail Transfer Protocol
SPIR	Symmetrically Private Information Retrieval
SR	Système de Recommandation
SSH	Secure SHell
SSL	Secure Sockets Layer
STPC	Secure Two Party Computation
TCP/IP	Transport Control Protocol/Internet Protocol
TTP	Trusted Third Party
URL	Uniform Ressource Locator
UF	Utility-Based Filtering
VPN	Virtual Private Networks
W3C	World Wide Web Consortium

à

mon épouse Ruth Nadège

mes enfants Marie Diane, Pierre-Ralph et Alfred Guerdy

REMERCIEMENTS

Je tiens à exprimer ici mes sincères remerciements à mes directeurs, la professeure Esma Aïmeur et le professeur Gilles Brassard, qui n'ont menagé aucun effort pour m'offrir un cadre de travail exceptionnel. Ils ont su veiller au bon grain pendant le processus, ô combien long et plein de défis, qui a abouti à la rédaction de cette thèse.

Je voudrais aussi remercier tous mes collègues étudiants et les professeurs des laboratoires HÉRON et LITQ pour leurs critiques et commentaires constructifs concernant mes travaux de recherche. Je dis particulièrement merci à Sébastien Gambs avec qui j'ai beaucoup discuté et échangé durant ma thèse.

Mes remerciements vont également à tous mes amis, en particulier Christophe Ewodo, Jean Paul Lafrenière et Rita Bélanger, ainsi que Maryam Erfani et sa petite famille, pour tous les services inestimables qu'ils m'ont rendus et qui ont contribué d'une certaine façon à l'aboutissement de cette thèse.

Je voudrais ensuite remercier mes proches qui ont suivi l'évolution de mon travail, malgré la grande distance qui me sépare de certains d'entre eux.

Pour finir, je remercie Nadège, Pierre-Ralph et Alfred, pour avoir pardonné ma nervosité circonstancielle et mes absences répétées à la maison...

Que chacun(e) trouve ici l'expression de ma très profonde gratitude !

INTRODUCTION

Les Technologies de l'Information et de la Communication (TIC) jouent un rôle très important dans les nombreux changements que connaît le monde. De nos jours, les disciplines qui ne sont pas influencées, d'une manière ou d'une autre, par les TIC sont plutôt rares.

Le commerce électronique ou e-commerce (*electronic commerce*) est l'une des disciplines qui prend appui sur les TIC. C'est un processus d'achat, de vente ou d'échange de biens et de services par l'intermédiaire d'un réseau de communication tel que l'Internet. De manière générale, deux entités, appelées *Client* et *Vendeur*, prennent part aux transactions commerciales en e-commerce. Par exemple, le vendeur propose des produits (biens et services) au client qui, en retour, les achète. Pour mieux servir le client, le vendeur a souvent besoin de colliger de l'information sur celui-ci pour en faire un profil.

De nos jours, de nombreuses technologies permettent de colliger des données sur les clients. Des simples *formulaires de saisie*, aux *cookies*, en passant par des *spywares*, *Web bugs*, *Log files*, etc. Ces technologies opèrent souvent de façon visible ou non pour le client. Ce dernier se retrouve donc au centre d'un processus dans lequel le vendeur se sert de ses données personnelles pour mener ses affaires. En particulier, le vendeur peut utiliser ces données pour faire ses propres statistiques, établir le comportement d'achat du client (qu'achète-t-il, où, en utilisant quel mode paiement ?), le retracer sur Internet, etc. Tout ceci peut se faire sans le consentement du client, ce qui constitue de manière évidente une violation de sa vie privée.

La vie privée est le fait pour un individu de faire une distinction entre les données qu'il juge publiques et celles qu'il considère comme étant privées. Si les premières peuvent être communiquées sans lui nuire, les secondes constituent plutôt des secrets à garder pour soi ou à partager avec quelques individus et organisations en qui il a *confiance*. Cette deuxième catégorie définit donc la sphère privée des données sur un individu, c'est-à-dire des données qui lui sont *personnelles*.

Face aux abus des vendeurs sans scrupules qui pourraient vendre, échanger ou partager des données personnelles sur les clients avec d'autres vendeurs, organisations et/ou gouvernements, sans l'aval de ces clients, il est utile de mettre sur pied des moyens de lutte contre toute forme de violation de la vie privée. Des lois ont ainsi été créées aux seins des gouvernements pour réglémenter l'utilisation des données personnelles colligées et gérées de manière électronique. De nombreuses organisations indépendantes ont aussi vu le jour dans le but de dénoncer tout abus constaté chez des vendeurs et surtout de d'informer de manière continue des individus. Seulement, si dans le cas des gouvernements, on ne saurait être à la fois juge et partie, du côté des organisations, il est difficile de contraintre les vendeurs à abandonner l'utilisation illicite des données personnelles sur des individus.

Dans cette thèse, nous introduisons une approche visant la protection de la vie privée des clients en e-commerce, tout le long du cycle d'une transaction d'achat. Notre approche vise aussi la protection des données sensibles des vendeurs qui sont souvent victimes de nombreuses attaques perpétrées contre leurs sites Web commerciaux. Nous avons ainsi développé un ensemble de protocoles pour réglementer de manière électronique, l'interaction entre le client et le vendeur. Pour protéger la vie privée du client et les données sensibles du vendeur, nous avons décidé de *masquer* les requêtes et réponses entre le client et le vendeur. Les protocoles développés reposent essentiellement sur les primitives et systèmes cryptographiques.

Cette thèse est organisé de la manière suivante : au chapitre 1, nous donnons un état de l'art du e-commerce. Le chapitre 2 est consacré à la cryptographie. Nous présentons ensuite la problématique liée à la vie privée au chapitre 3. Nous introduisons notre approche au chapitre 4 et détaillons les protoles d'achat masqué au chapitre 5.

CHAPITRE 1

COMMERCE ÉLECTRONIQUE

L'informatique est au service de l'esprit créatif de l'Homme qui lui trouve chaque jour de nouvelles fonctionnalités. Cependant comme toute invention, elle sert également les esprits détraqués et malicieux

D. Godart

1.1 Introduction

Le commerce est un mécanisme qui permet à deux parties, le vendeur et le client, de mener des activités de vente, d'achat et/ou d'échange de produits (biens, services ou informations), sur la base de la théorie de l'offre et la demande. Contrairement à sa forme traditionnelle dans laquelle le vendeur, le client et éventuellement un ou plusieurs intermédiaires effectuent *physiquement* une transaction de vente ou d'achat, le commerce s'est transporté, depuis quelques décennies, dans un monde virtuel dominé par les réseaux informatiques, l'Internet en l'occurrence, pour donner naissance au *Commerce Électronique* [176] ou *e-commerce*. Dans ce contexte virtuel, les deux entités peuvent ne jamais se rencontrer et/ou reconnaître avoir effectué des transactions électroniques. Après l'achat, la livraison peut être physique ou électronique, tout dépend du bien ou du service acheté. Le commerce électronique est un marché planétaire. Selon ses capacités, on y accède par l'intermédiaire de l'électronique et parfois physiquement en complément. Le commerce traditionnel a des règles qui évoluent rarement. Il reste bien solide jusqu'ici et bénéficie d'une grande confiance chez ses participants. En revanche, le commerce électronique est un phénomène très récent qui se construit chaque jour davantage. Il suit un chemin plein d'obstacles et essaie autant que faire se peut de les contourner.

Dans ce chapitre, nous faisons un survol du commerce électronique. En partant de son historique (section 1.2), nous allons tour à tour présenter ses fondements (section 1.3), ses différents types (section 1.4), le modèle CBB (*Customer Buying Behaviour*) [91] (section 1.5) et les systèmes de recommandation (section 1.6).

1.2 Historique

Les premières applications du commerce électronique ont vu le jour dans les années 1970 à travers des initiatives tel que le transfert électronique de fonds ou EFT (Electronic Fund Transfers) [141]. Toutefois, l'exploitation de ces applications était limitée aux grandes entreprises, aux institutions financières et à quelques petites entreprises audacieuses [176]. Par la suite, les échanges de données informatisées ou EDI (Electronic Data Interchange) [2] voient le jour et surpassent les simples transactions financières. Les EDI vont de ce fait augmenter le nombre de compagnies participantes, allant des institutions financières aux fabricants, en passant par les détaillants, les sociétés de services, et beaucoup d'autres types d'entreprises. De plus, de nouvelles applications du commerce électronique vont naître : gestion de stocks, systèmes de réservations pour les voyages, etc.

Avec la grande commercialisation de l'Internet et la participation accrue des internautes au *World Wide Web* dans les années 90, le commerce électronique s'est affirmé davantage et ses applications se sont grandement répandues. Parmi les raisons qui ont provoqué cette rapide expansion [157], on peut citer la mise sur pied de nouveaux réseaux, protocoles et logiciels et le développement de nouvelles spécifications à l'instar des applications dédiées Internet (pages web dynamiques, téléchargement, etc.). Le poids de la concurrence et d'autres pressions dans les affaires sont aussi des raisons non négligeables. Les pressions dont il est question ici peuvent être d'ordre social, économique, législatif, technologique ou politique ; étant entendu que tous ces secteurs ont une certaine influence sur les activités du monde des affaires.

Depuis 1995, les utilisateurs d'Internet sont témoins du développement de plusieurs applications telles que les publicités interactives, les expérimentations virtuelles, etc. Presque toute compagnie, quelle que soit sa taille, dispose désormais d'un site Web et, au Canada et aux États-Unis par exemple, la plupart des grandes compagnies ont des portails électroniques [176].

De nos jours, les statistiques sur la taille future du commerce électronique sont très changeantes et dépendent largement des pays. Au Canada par exemple, les achats via Internet atteindront 4,6 milliards de dollars US pour l'année 2005 contre 142,5 milliards de dollars US pour les États-Unis. Ces statistiques sont aussi fonction des habitudes de consommation des habitants des différents pays. Pour l'année 2004, les dépenses en ligne annuelles par habitant du Canada et des États-Unis ont été respectivement de 117 et 692 dollars US. Par ailleurs, le nombre d'internautes attendus en 2008 dans le monde entier est de 750 millions et les experts (`forrester.com`) prédisent qu'à peu près 50% de tous les utilisateurs d'Internet vont faire des achats en ligne. Des données plus récentes sur les statistiques en commerce électronique sont accessibles à travers les sites Internet `forrester.com`, `eMarketer.com`, etc.

Après ce bref parcours de l'historique, nous allons à présent parler des fondements mêmes du commerce électronique. Cela nous permettra d'avoir une vue globale de ce processus électronique.

1.3 Fondements du commerce électronique

Dans cette section, nous allons couvrir les éléments clés du commerce électronique. Nous commencerons par la définition de quelques concepts, puis nous poursuivrons avec les entités, les différents mécanismes, les disciplines, les pressions, les bénéfices et, pour finir, les limites.

1.3.1 Définitions

Réseau informatique : Un réseau informatique est un ensemble d'ordinateurs et d'autres appareils périphériques (imprimantes, numériseurs, etc.) interconnectés. Le commerce électronique prend appui sur de tels réseaux pour se développer. Dans cette section, nous présentons essentiellement trois types de réseaux informatiques : Internet, Intranet, Extranet.

Internet : Développé à l'origine par le ministère de la Défense des États-Unis [1], l'**Internet** (**Inter**connected **net**works) est de nos jours le plus grand réseau informatique (ensemble d'ordinateurs interconnectés) du monde. C'est un réseau planétaire auquel sont raccordés les ordinateurs de compagnies, d'organismes gouvernementaux, d'établissements d'enseignement, de centres hospitaliers, etc. Il repose sur un système universel d'adressage utilisant le protocole de communication TCP/IP (*Transmission Control Protocol/Internet Protocol*) pour permettre à ses millions d'utilisateurs d'avoir accès aux ressources disponibles tels que le courrier électronique, les protocoles de transfert de fichiers, les nouvelles, le navigation web (Browsers), entre autres. L'apport le plus intéressant d'Internet réside en sa capacité à s'articuler autour d'un protocole de communication indépendant du type de machine (PC, Mac, etc.), du système d'exploitation (Linux, Windows, MacOS, etc.) et du support de transport physique utilisé.

Intranet : L'Intranet est un réseau local ou LAN (*Local Area Network*) appartenant à une entité (compagnie, organisme gouvernemental, établissement d'enseignement, etc.). L'accès à un tel réseau est souvent très limité : des procédures d'accès sont mises en places en fonction de la qualité des individus (membres ou employés, clients, fournisseurs, etc., de l'entité considérée). Parfois, l'Intranet se limite simplement à ce qui est palpable, le site web interne à la compagnie par exemple. L'Intranet utilise les mêmes technologies que celles d'Internet. On parlera ainsi des concepts clients et serveurs qui communiquent par le biais du protocole TCP/IP. Dans le même sens, les protocoles de navigation sur Internet, de transfert de fichiers, de courrier électronique, etc., sont aussi communément utilisés dans un Intranet.

L'accès à Internet à partir d'un Intranet n'est pas souvent immédiat. Il se fait généralement à travers un coupe-feu (*firewall*) qui sert de point de contrôle pour la sécurité de l'Intranet.

Firewall : Un Firewall est un ensemble de mécanismes permettant de partitionner un réseau ou sous-réseau afin de protéger une partie, normalement interne ou privée, de l'autre partie, externe ou publique, moyennant l'application d'une politique de sécurité.

On peut classifier les firewalls en deux catégories principales : les filtres de paquets (*packet filters*) et les passerelles (*gateways*).

Les filtres de paquets sont des composants permettant de bloquer le trafic entre deux réseaux suivant un ensemble de règles. Les passerelles sont des composants permettant de relayer le trafic entre deux réseaux ou deux sous-réseaux ; ils peuvent intervenir au niveau de la couche transport ou au niveau de la couche application du modèle TCP/IP [134].

Filtrage de paquets : Le filtrage de paquets (*packet filtering* ou *screening*) est une technique [134] permettant de contrôler le flot de paquets suivant un ensemble de critères entre deux réseaux ou sous-réseaux. Il constitue souvent une première barrière, simple à mettre en place et très efficace pour protéger un Intranet (réseau interne à une compagnie) des attaques externes. Elle rend également possible le cloisonnement d'un Intranet en domaines de sécurité distincts (par exemple, la sécurité dans les laboratoires de recherche). La politique de base du filtrage est que tout ce qui n'est pas explicitement autorisé est interdit. Naturellement, le filtrage de paquets se sert des filtres de paquets que nous venons de présenter.

Extranet : L'Extranet est un réseau informatique qui utilise l'Internet pour relier deux ou plusieurs réseaux Intranet. C'est en fait une extension du système d'information d'une compagnie à des partenaires ou à des filiales situés au-delà de son réseau Intranet. Une telle extension signifie un accès privilégié à certaines ressources informatiques de la compagnie par l'intermédiaire d'une interface Web (Internet).

Commerce électronique : L'évolution du commerce électronique fait que sa définition tente chaque jour de s'adapter au nouveau contexte qui se présente. Pour nous, le commerce électronique est un processus d'achat, de vente, de transfert et d'échange de biens, de services et/ou d'informations à travers un réseau d'ordinateurs, en l'occurrence l'Internet, ou tout autre moyen de communications (ex. : le réseau mobile). `Ebay.com`, `amazon.com`, `entrust.com` sont des exemples de sites Web qui pratiquent le commerce électronique.

De façon plus restreinte, la définition du commerce électronique est fonction du contexte mis en jeu (communications, affaires ou services) [161]. En rapport avec les communications, le commerce électronique se réduit à la livraison des biens (cas des logiciels), des services (ex. : le support technique), des informations (ex. : les adresses) et des paiements au moyen des réseaux d'ordinateurs ou par un quelconque moyen électronique.

Pour ce qui est du monde des affaires, le commerce électronique conduit à l'automatisation des procédures (ventes par exemple) et des transactions de préférence commerciales, sans oublier qu'il influence la chaîne de travail avec, par exemple, la réduction de la main d'oeuvre. Le commerce électronique est par ailleurs un outil qui permet aux compagnies (de vente), aux clients et à l'administration de réduire les coûts des services tout en garantissant (d'une manière générale) une meilleure qualité et un court délai de livraison. On parle de services en ligne pour dire que le commerce électronique offre la possibilité de faire des achats et des ventes de produits et d'informations sur Internet.

Par ailleurs, le commerce électronique facilite la collaboration [176], avec des regroupements inter et intra-compagnies, ainsi que la création des communautés virtuelles ; par exemple, la prolifération d'agences matrimoniales sur Internet.

Boutique virtuelle : Nous définissons une boutique *traditionnelle* comme étant un lieu physique où l'on achète des biens et/ou des services. Les super marchés, les magasins de ventes et autres centres d'achats, constituent des exemples de boutiques traditionnelles.

Une boutique virtuelle est en revanche un espace réservé d'un site Web à travers lequel le vendeur *entrepose* des biens, services et informations qu'il vend pour permettre à l'internaute (client potentiel) de les identifier, commander et acheter en ligne. D'une certaine façon, la boutique *virtuelle* est *la* version électronique d'une boutique traditionnelle, les ventes, achats et échanges étant simplement faits à travers un site Internet.

Catalogue électronique : Un catalogue *électronique* est une description à travers des pages Web des biens et services vendus sur l'Internet. Il porte aussi le nom de *e-catalog*.

D'une manière générale, il peut se résumer en une base de données de tous les produits (biens et services) que le vendeur met à la disposition de ses clients. Les catalogues électroniques peuvent être regroupés suivant trois dimensions :

- le dynamisme de la présentation du contenu : ceci s'oppose à la version papier du catalogue (version traditionnelle), donc statique ;
- le degré de personnalisation : c'est la possibilité pour le catalogue du vendeur d'adapter son contenu au profil du client (section 1.6.2) ;
- l'intégration dans le processus d'affaire : le catalogue doit tenir compte du modèle d'affaire mis en place par le site de commerce électronique.

Portail électronique : Un *portail électronique* ou simplement *portail* est un site central à partir duquel l'on peut retrouver des informations sur des produits et autres services [115,147]. Le lien n'est pas direct. Le portail ne sert qu'à rediriger l'utilisateur vers les sites qui vendent effectivement les produits ou les services. En pratique, un portail électronique joue en quelque sorte un rôle de porte d'entrée de plusieurs boutiques virtuelles dans le commerce électronique, ce qui signifie que c'est la porte d'entrée d'un certain nombre de boutiques sur Internet. Par ailleurs, à partir d'un portail, on peut récolter des revenus en rapport avec les publicités interactives, les liens vers les sites de vente, etc. `Yahoo.com`, `Amazon.com`, `Altavista.com`, etc., constituent des exemples de portails électroniques. Un portail électronique peut aussi être un site central qui donne l'information sur tous les services offerts par une entité donnée, un gouvernement par exemple.

Après ces définitions, nous allons nous tourner vers les différentes entités qui participent au développement et à la pratique du commerce électronique.

1.3.2 Entités

Les acteurs du commerce électronique font partie d'un espace de marché (*Marketspace*) virtuel, qui est le lieu de rencontre des vendeurs et des clients pour échanger des produits contre l'argent ou d'autres produits. Un espace de marché se compose des éléments suivants [176] : les vendeurs, les clients, les produits, l'infrastruture, l'interface client, l'interface vendeur, les intermédiaires, les partenaires d'affaire et les services de support.

- Le ***vendeur*** est un individu ou groupe d'individus (une organisation par exemple) détenteurs de produits (biens, services et informations) à vendre. La description de ces produits ainsi que leurs prix et conditions de vente sont mis ensemble pour former le catalogue électronique du vendeur.
- Le ***client*** est un individu ou groupe d'individus (une organisation par exemple) à la recherche d'un produit. Parfois, il connaît juste son besoin et n'a pas nécessairement une idée précise du produit qui pourrait le satisfaire. Dans tous les cas, le client va interroger le catalogue du vendeur pour rechercher un produit précis (s'il le connaît) ou alors pour décrire son besoin et recevoir des propositions de produits, par l'intermédiaire d'un système de recommandation (section 1.6).
- Le terme ***produit*** utilisé dans le cadre de cette thèse renvoie à un bien, un service ou une information que le client achète auprès du vendeur. Évidemment, il faut tenir compte du contexte dans lequel il est utilisé pour lui donner le sens que nous venons de mentionner.
- L'***infrastructure*** est l'ensemble des moyens matériels utilisés en commerce électronique ; par exemple, les ordinateurs, les réseaux informatiques et de télécommunications, etc.
- L'***interface client*** ou *front-end* est un ensemble de programmes dits d'interface à travers lesquels le client interagit avec le système du vendeur. Le portail électronique, le catalogue électronique, le panier d'achats, le moteur de recherche et les passerelles vers les systèmes de paiement sont des exemples de programmes que le vendeur peut mettre dans l'interface client.
- L'***interface vendeur*** ou *back-end* met en relief les différentes activités qui permettent de gérer en ligne la passation de commandes de produits. Cette composante inclut donc, entre autres, la gestion d'inventaires, les achats auprès des fournisseurs, le traitement des paiements, l'emballage et la livraison.
- Un ***intermédiaire*** en commerce électronique est une entité physique ou virtuelle qui assiste le client et le vendeur à finaliser un processus de vente/achat. Les

banques (pour les paiements électroniques), les compagnies de livraison de produits, les agents mobiles, entre autres, constituent des exemples d'intermédiaires en e-commmerce. De façon générale, les intermédiaires fournissent des services à valeur ajoutée aux vendeurs et aux clients. Le terme *Infomédiaire* désigne les intermédiaires électroniques, ceux-là même qui cherchent à contrôler la circulation de l'information sur Internet, et qui peuvent en créer des agrégats à vendre. Les boutiques du coin (*convenient stores*) et les compagnies de publicité sont d'autres exemples d'intermédiaires. Les premières constituent parfois des points de livraison des produits physiques achetés en commerce électronique. Quant aux secondes, elles se chargent de promouvoir des produits à la demande des vendeurs (ex. : `doubleclick.com`). Un dernier exemple d'intermédiaire est relatif à des compagnies organisatrices des conférences. De telles compagnies font tout ce qui est nécessaire à la réussite des conférences, à l'exception de l'évaluation des articles qui y sont soumis.

- Un ***partenaire d'affaire*** d'une compagnie est une personne physique ou morale qui est impliquée dans sa gestion, en tant qu'associé. Le partenaire d'affaire peut par exemple disposer des parts dans la compagnie. Dans le cadre du commerce électronique, le vendeur peut s'associer à des agences de livraison pour faciliter l'acheminement des produits à ses clients. Il peut aussi signer des contrats de représentabilité dans le souci de couvrir les zones qui lui sont difficiles d'accès, à cause par exemple de l'éloignement.
- Un ***service de support*** est généralement apporté au client, sous l'appellation *support technique*, pour gérer l'après-achat. Plus précisément, cette composante permet au vendeur d'apporter une assistance continue à ses clients pour ce qui est de la maintenance et des éventuelles mesures correctives applicables aux produits vendus. Le service de support est un autre exemple d'élément pouvant faire l'objet de partenariat ; c'est-à-dire, le vendeur confie la gestion de son service après-vente [1] à une compagnie tierce.

[1]On parle de "service après-achat" du point de vue du client

La pratique du commerce électronique consiste essentiellement en un certain nombre de mécanismes. C'est de ces mécanismes que nous allons parler dans la section suivante.

1.3.3 Mécanismes

Le commerce électronique s'intéresse à la vente des biens matériels de toutes sortes, à la commercialisation des services (cas des services bancaires avec la consultation des comptes, les virements, etc.), à la diffusion des produits immatériels (téléchargement des logiciels, des jeux vidéo, etc.). Par ailleurs, une pratique efficace du commerce électronique suppose la mise sur pied d'un certain nombre de fonctionnalités : l'acquisition et le stockage des données, l'extraction des données, les paiements électroniques, la sécurité des services, la connectique (voir mécanisme de connexion, plus bas), les exigences légales, etc. La présente section décrit donc quelques mécanismes relatifs à ces intérêts et fonctionnalités.

1.3.3.1 Mécanismes de ventes

Pour vendre, plusieurs mécanismes sont mis en œuvre sur Internet [83]. Parmi les plus connus, nous avons : les bons de commandes pour l'achat d'un produit spécifique, les paniers de commandes (ou paniers d'achat) pour l'achat simultané de plusieurs produits, les systèmes de réservation (billets d'avion, location de véhicules, chambres d'hôtels), les abonnements à des services d'information (presses, magazines), les ventes aux enchères, etc. Les mécanismes d'échanges peuvent aussi être répertoriés ici, c'est-à-dire un produit ou un service est échangé contre un autre produit ou service.

1.3.3.2 Mécanismes d'achats

L'application qui gère les achats [147] est généralement stockée dans un serveur sécurisé. Nous considérons le cas typique d'un achat sur Internet : le client agit depuis son domicile à l'aide d'un ordinateur. Quand un tel client prend la décision de faire un achat en ligne, il est invité par le serveur sécurisé à remplir un formulaire avec un certain nombre d'informations, en particulier, les produits désirés, son adresse et son mode de paiement (carte de crédit par exemple). Par la suite, il valide la transaction et reçoit à l'écran un message lui signifiant que la transaction a été effectuée avec succès et, dans le même sens, un e-mail (electronic mail) de confirmation suit pour le remercier et lui fournir diverses informations utiles au suivi de sa commande. Les données de la transaction sont quant à elles stockées dans un serveur de base de données et les services de commandes/livraisons et/ou de comptabilité (produits matériels ou immatériels, services) peuvent ainsi prendre la relève pour faire soit une expédition du (des) produit(s) acheté(s), soit prendre simplement connaissance du nouvel achat.

1.3.3.3 Mécanisme d'acquisition et de stockage des données

L'acquisition et le stockage des données [2] sont liés au problème de la numérisation des informations disponibles (catalogues de produits, livres, films, cartes géographiques, etc.). Ce problème met en relief trois grandes préoccupations. *D'abord*, les données doivent être mises dans un format spécifique afin de rendre leur utilisation facile. *Ensuite*, il faut tenir compte de l'impact économique, étant donné que les coûts de la numérisation de l'information sont parfois très élevés. *Enfin*, il faut disposer des moyens logiciels et matériels permettant d'acquérir et de stocker les données au fur et à mesure que les sources d'informations (particuliers, compagnies, corporations, etc.) les rendent disponibles. Les moyens logiciels et matériels peuvent par exemple aider à transformer les données provenant de diverses sources en blocs d'une certaine taille, faire l'analyse de ces blocs et charger leurs contenus dans des bases de données, faire la mise à jour des bases de données, etc. Une fois les données acquises et stockées, l'utilisateur (le client par exemple) doit être en mesure des les extraire.

1.3.3.4 Mécanisme d'extraction des données

Les besoins en extraction de données les plus connus en commerce électronique concernent les catalogues électroniques, le filtrage des données (on se limite autant que possible aux données correspondant à la requête du client), les moteurs de recherches (`Yahoo.com`, `Altavista.com`, `Google.com`, etc.), les *agents* (programmes appelés à exécuter une fonction spécifique sur Internet à la place d'un humain). Il est aussi important de prendre en compte les nouveaux types de données tels que les enregistrements multimédia (vidéo, son, images 3D) ; de développer de nouvelles méthodes permettant d'indexer les données ; de s'intéresser à la négociation des agents entre eux [17, 91, etc.]) ; etc.

1.3.3.5 Mécanisme de paiement

Pour acheter ou pour vendre des biens ou des services sur Internet, il faut se servir d'un certain nombre de moyens de paiements : carte de débit, carte de crédit, chèque électronique (*e-check*), monnaie électronique (*e-money*), etc. Le chèque et la monnaie électroniques font appel à une simulation électronique des échanges faits dans la vie courante. Dans cette simulation, la monnaie et le chèque électroniques sont généralement gérés par une tierce partie qui procède au paiement effectif des sommes dues. Pour être titulaire d'un chèque ou d'une monnaie électronique, l'une des situations les plus courantes est de faire au préalable un dépôt auprès de la tierce partie (`Paypal.com` par exemple) qui donne alors un numéro d'identification de dépôt à utiliser dans les différentes transactions. Après un achat, le compte correspondant au numéro d'identification est débité du montant de la transaction. De façon particulière, un chèque électronique requiert un service d'authentification (chapitre 2) pour mieux distribuer les informations entre le client (payeur), le vendeur (payé) et les banques. Plusieurs technologies sont mises en œuvre pour supporter les paiements électroniques en général, la monnaie et le chèque électroniques en particulier [83, 176].

1.3.3.6 Mécanisme de sécurité

Il faut sécuriser les systèmes de commerce électronique contre les intrusions internes et/ou externes. Nous parlerons des différents services de sécurité nécessaires à la pratique du commerce électronique au chapitre 2.

1.3.3.7 Mécanisme de connexion

De nos jours, l'environnement du commerce électronique fait face à un grand nombre de clients et de vendeurs. Les vendeurs disposent de boutiques sur Internet et les clients y accèdent en se servant des ordinateurs de bureau, des ordinateurs portables, des téléphones fixes et mobiles, des téléviseurs, etc. La connectique [99, 134, 168] renvoie donc à tous les moyens dont se servent les clients ou les vendeurs pour avoir accès à l'environnement du commerce électronique.

1.3.3.8 Mécanisme de législation

L'aspect légal dans le commerce électronique est très préoccupant. En effet, les problèmes liés à la propriété intellectuelle, à l'imposition, à l'application de la loi, à l'omniprésence de la cryptographie, au contenu et au respect des contrats n'ont toujours pas de solutions définitives dans l'industrie croissante du commerce électronique. Par exemple, il y a lieu de se demander quelle législation appliquer pour un citoyen français, ayant un fournisseur d'accès Internet allemand, qui fait ses achats auprès d'une compagnie canadienne via son site Web aux États-Unis, et qui paie avec sa carte de crédit émise au Royaume-Uni ? De nos jours, on cherche sans cesse à rapprocher les différentes législations [2, 200].

Un autre aspect de la législation, le plus important pour nous, concerne la vie privée. Nous y reviendrons au chapitre 3.

Le commerce électronique est un nouveau champ d'investigation pour les chercheurs et ses fondations théorique et scientifique deviennent plus solides chaque jour. Les disciplines qui composent le commerce électronique sont nombreuses et ci-dessous, nous ne nous intéressons qu'à une infime partie d'entre elles.

1.3.4 Disciplines

La mise sur pied et l'évolution quotidienne du commerce électronique font appel à plusieurs disciplines [176]. Nous en listons quelques uns dans la suite de cette section.

L'informatique : Le commerce électronique intéresse tous les domaines de l'informatique : langages de programmation, réseaux, bases de données, multimédia, intelligence artificielle, etc. L'informatique constitue d'ailleurs le maillon central autour duquel gravitent les autres disciplines qui influencent ou sont influencées (par) le commerce électronique.

Le marketing : Bien que les besoins du client ne soient pas toujours fonction des produits existants, pour vendre, le bien ou le service concerné doit être connu du client ; aussi faut-il trouver un moyen efficace pour attirer l'attention de ce dernier. Les publicités et les stratégies qui les accompagnent (le *spam* par exemple), les kiosques interactifs, etc., constituent quelques éléments clés du marketing dans la pratique du commerce électronique.

Les finances : L'importance des institutions financières et des banques n'est plus à démontrer quand il s'agit de faire du commerce. Dans le cadre du commerce électronique, ces deux entités se retrouvent au centre de toute transaction financière en ligne, ce qui suppose qu'elles subissent davantage les attaques des groupes du crime organisé et autres malfaiteurs. Il faut donc trouver des stratégies très efficaces pour faire face à ces manœuvres qui peuvent causer, par exemple, la perte d'importantes sommes d'argent.

L'économie : Le commerce électronique est influencé par les différentes forces économiques et a un impact majeur sur les économies nationales et l'économie mondiale. Par exemple, les transactions en ligne sont génératrices de revenus. D'une manière générale, les économistes examinent constamment comment le commerce électronique agit sur les compagnies et les corporations.

La gestion : De nouvelles approches et théories dans la gestion doivent être mises en place compte tenu de la nature interdisciplinaire du commerce électronique. Par exemple, une compagnie qui ouvre une boutique virtuelle ne devrait plus se limiter simplement à la production d'un bien. Elle devrait plutôt revoir ses services de marketing, d'informatique, de gestion de la clientèle (CRM ou *Customer Relation Management*)... et même la formation de ses employés.

La comptabilité : Des transactions off-line qui ne nécessitent pas la connexion sur Internet sont faites dans le souci d'auditer les systèmes informatiques et de générer des écritures (comptables) exploitables par le service de comptabilité. Même les transactions sur Internet simplifient les tâches de comptabilité grâce à la génération automatique d'écritures comptables dans une base de données du vendeur. De nos jours, les bons de commandes papier prennent de plus en plus du large au bénéfice des commandes electronique.

La cryptographie : Son importance est naturelle dès lors qu'il faut sécuriser les transactions en ligne. Les paiements et l'intégrité des données sur Internet (ou tout autre réseau d'ordinateurs) ou dans les serveurs de données en dépendent fortement. De nombreux protocoles cryptographiques dédiés au commerce électronique ont déjà vu le jour. Nous y reviendrons aux chapitres 2 et 3.

Le droit : Il est question, entre autres, de protéger la vie privée, de gérer les droits d'auteurs et d'assurer la protection de la propriété intellectuelle dans un réseau ouvert. Dans le même sens, il faut protéger les utilisateurs sans défense ou peu avertis, comme c'est le cas des enfants qui se retrouvent face à des sites Web dont le contenu a peu de crédit ou trangresse les valeurs éthiques (films pour adultes, jeux de hasard, etc.).

La liste ci-dessus est loin d'être exhaustive car toutes les disciplines de divers domaines, scientifiques ou non, participent dans une certaine mesure à la mise en œuvre du commerce électronique. Grâce à ces disciplines, plusieurs types de commerce électronique sont quotidiennement créés et pratiqués. Avant de les présenter, nous allons d'abord donner un bref aperçu des pressions que subit le commerce électronique dans sa pratique, ainsi que les bénéfices qu'il apporte aux compagnies, aux clients et, d'une manière générale, aux populations.

1.3.5 Pressions

Le commerce électronique fait face à une triple pression issue de l'environnement des affaires [176].

Pression du marché et de l'économie : Il faut tenir compte de la forte compétition issue d'une économie qui se veut mondiale, de la constitution des partenariats régionaux, du pouvoir d'achat très insignifiant dans certains pays, des changements fréquents et significatifs dans les marchés et des consommateurs qui deviennent davantage exigeants et puissants face à la pluralité des offres.

Pression sociale et environnementale : Il est question de changer les méthodes de travail et de tenir compte des lois et règlements des gouvernements, des aides aux gouvernements, des droits et devoirs des populations, des éventuels changements politiques, etc.

Pression technologique : Une technologie peut très vite connaître une obsolescence ; aussi faut-il tenir compte des grandes innovations et des nouvelles technologies, de la désinformation et du déclin rapide des coûts des technologies face à leur côté performance.

Face à ces pressions, les compagnies adoptent des solutions pour se maintenir. On peut citer, entre autres :

- *Les systèmes stratégiques :* les compagnies mettent l'accent sur le partage du marché, la meilleure négociation avec leurs fournisseurs et évitent que les concurrents entrent dans leurs territoires de prédilection.
- *Le renforcement continu des efforts :* la plupart des compagnies optent pour la continuité de leurs programmes d'amélioration de la production, de la qualité et du service aux clients. Ces compagnies versent également dans la créativité et les

changements dans leurs administrations.

- *La ré-ingénierie des processus de traitements :* dans ce contexte, on revoit la périodicité des livraisons, la collaboration dans le travail ; on opte pour la vente en masse, la restructuration des équipes de base, etc.
- *Les alliances d'affaires :* bon nombre de compagnies réalisent que des alliances avec d'autres compagnies, même concurrentes, peuvent leur générer beaucoup de bénéfices.

1.3.6 Bénéfices

Rares sont les innovations dans l'histoire de l'homme qui ont facilement engendré un potentiel de bénéfices tel que le commerce électronique le fait [176]. Tout le monde semble y aller gagnant. Compagnies et particuliers y trouvent leur compte. Tout ceci est dû à la nature globale de la technologie utilisée dans le commerce électronique, à sa politique d'interpellation de centaines de millions d'individus, à son interactivité, au nombre élevé de possibilités en vue de son utilisation, à son grand nombre de ressources et à la rapide expansion des infrastructures qui le supportent, spécialement le Web. Les bénéfices restent quelque peu freinés par les obstacles (sécurité, sites de vente fictifs, hésitations de la part des clients, etc.) [148, 176] rencontrés depuis sa naissance. Mais, compte tenu du fait que ces obstacles diminuent progressivement si on s'en tient aux nombreux développements quotidiennement faits, il est clair que les bénéfices qui ont commencé à se matérialiser, vont connaître une augmentation significative au fur et à mesure de l'expansion de leur générateur (commerce électronique). C'est d'ailleurs ce qui a fait dire à Bill Clinton et Al Gore, en 1997 [63], que la révolution du commerce électronique est juste « aussi profonde que le changement qui vint avec la révolution industrielle ».

En réalité, les bénéfices sont fonction de l'entité concernée, selon qu'il s'agit d'une compagnie, d'un client ou de la société tout entière. Commençons par les compagnies.

Bénéfices pour les compagnies : Loin d'avoir une liste exhaustive des bénéfices, nous ne parlons ci-dessous que de ceux qui nous semblent les plus significatifs.

Parmi les bénéfices, on peut citer l'expansion des places pour les marchés nationaux et internationaux, car avec un capital minimal, une compagnie peut facilement et rapidement localiser un grand nombre de clients, les meilleurs fournisseurs et les partenaires d'affaires les plus sûrs à travers Internet. Comme autres avantages, il y a la réduction des coûts de fabrication, de traitement, de distribution, de stockage et de fouille dans l'archivage papier ; la diminution des coûts d'administration ; la minimisation de l'inefficacité de la chaîne d'approvisionnements (par exemple [176], la fourniture des services par le biais de l'électronique peut conduire à la réduction des coûts d'administration pour plus de 80%, réduisant par la même occasion les prix d'achats de 5 à 15% et les délais de livraison de plus de 50%).

Par ailleurs, les compagnies traitent de manière plus intime avec les clients. L'Internet rapproche la compagnie du client puisque ce dernier a accès à l'information désirée depuis son domicile par exemple. Pour les compagnies, la désintermédiation (le fait d'éliminer la présence des intermédiaires entre la compagnie et ses clients) prend de plus en plus place. Par exemple, les intermédiaires "temps" et "personnel" peuvent être éliminés par l'automatisation des procédures.

Dans le cadre du commerce électronique, la taille d'une compagnie n'est pas un facteur important (puisqu'elle reste généralement inconnue ou alors difficilement vérifiable), les compagnies de petite ou moyenne taille ont une chance presque égale de se faire une bonne clientèle ; étant donné que la localisation d'une compagnie n'a pas d'influence sur l'accès du client à son site, le support à la clientèle peut aller au-delà des barrières géographiques.

A ces bénéfices, nous pouvons ajouter des exemples [176] précis illustrant des différences importantes entre les coûts pour les ventes conventionnelles et ceux via l'Internet.

Pour commencer, il coûte à la banque 1,08$US pour conduire à terme une simple transaction au moyen d'un guichet automatique, alors que sur Internet, la même transaction ne coûte que 0,10$US. C'est d'ailleurs la raison pour laquelle plusieurs banques et institutions financières encouragent leurs clients à faire des transactions financières via Internet. Un autre exemple vient du coût d'émission d'un billet d'avion. Sur le Web, ce coût est de 1$US tandis que dans un système physique (agence de voyages), une telle transaction affiche environ 8$US. Pour continuer, il coûte au gouvernement américain 0,43$US pour émettre un chèque sur support papier contre 0,02$US pour un envoi électronique du même paiement. Par ailleurs, l'envoi d'une facture revient à 1,60$US contre une réduction de moitié, soit 0,80$US, pour les factures électroniques.

Bénéfices et avantages pour les clients : Comme premier élément important, les clients font des achats et d'autres transactions sur Internet 24 heures par jour et à tout moment dans l'année ; tout ceci à partir de n'importe quel point du globe terrestre (ce qui élimine les coûts de déplacement, entre autres), sous réserve de l'existence des infrastructures nécessaires. Ensuite, le commerce électronique offre plusieurs choix aux clients car, ils peuvent faire des sélections auprès de plusieurs vendeurs et à partir de nombreux produits. De même, les clients bénéficient d'une grande réduction des prix pour les produits et services puisqu'ils sont appelés à fouiller dans plusieurs boutiques et à faire une comparaison rapide des prix.

Notons aussi que pour les produits numériques, la livraison est quasiment immédiate et, d'une manière générale, les clients peuvent aisément obtenir le détail des produits sollicités en quelques secondes au lieu d'attendre des journées entières, voire des semaines. Par ailleurs, la possibilité de participer à des enchères virtuelles est offerte et peut être très bénéfique pour le client. Il court la chance d'acquérir des produits à des prix dérisoires. Pour finir, les clients peuvent se regrouper en communautés virtuelles dans le but d'échanger des idées et de partager les expériences vécues par les uns et les autres [83].

Bénéfices généraux pour les populations : Le travail à domicile et partant, le petit nombre de déplacements et la réduction du trafic sur les routes et celle de la pollution constituent des avantages inestimables pour les populations. De plus, les marchandises étant parfois vendues à très bas prix, la plupart des personnes peuvent enfin atteindre un niveau de vie acceptable.

Les populations du tiers monde peuvent désormais accéder à des produits et services qu'elles n'auraient presque jamais connus sans l'existence de l'Internet et du commerce électronique. De nos jours, ces populations parviennent à suivre des formations professionnelles et à obtenir des diplômes en ligne. Pour finir, les services publics tels que les services sociaux, la santé, l'éducation peuvent être offerts à des coûts très réduits et/ou avec une qualité supérieure. Par exemple, les médecins des zones rurales (ou très réculées) peuvent avoir accès à des informations et à des technologies de pointe pour mieux traiter leurs patients, dépendamment de la disponibilité de l'infrastructure adéquate (section 1.3.2).

Malgré tous ces bénéfices, le commerce électronique reste soumis à plusieurs limites qui freinent quelque peu son expansion.

1.3.7 Limites

Les limites du commerce électronique sont à la fois d'ordre technique et non technique.

Sur le plan technique, les systèmes de sécurité, la fiabilité, les standards et les protocoles de communication continuent de connaître des améliorations, ce qui peut causer des problèmes d'instabilité. Il faut noter que dans plusieurs régions, les bandes passantes sont très limitées. Par exemple, dans certains pays où l'Internet est servi aux usagers via la liaison téléphonique, même si ces usagers disposent de modems de connexion de grande capacité, le débit peut rester très limité (9,6 kbits/s dans certains pays du tiers monde par exemple). La gestion de l'information (ou des données) peut également constituer une limite technique sur les plans du stockage, de l'extraction et du cheminement de l'information à travers l'Internet. Il faut aussi noter que le développement des applications dédiées au commerce électronique vit des améliorations quotidiennes, ce qui a une certaine influence sur les usagers. De même, il est parfois difficile d'intégrer l'Internet et les applications de commerce électronique à certains systèmes (logiciels, bases de données) existants.

Par ailleurs, les vendeurs ont besoin de serveurs Web spéciaux, de serveurs de réseaux particuliers et d'autres infrastructures de développement pour opérer. Il arrive que les applications de commerce électronique soient incompatibles avec le matériel et/ou le système d'exploitation de l'usager, ce qui conduit à un nouvel investissement !

Ces limites techniques sont importantes, mais au fil du temps, elles sont de plus en plus minimisées.

Sur le plan non technique, nous pouvons citer plusieurs grandes préoccupations [176]. D'abord, dans le cas de la vente en détail en ligne, il faut particulièrement assurer la sécurité et la protection de la vie privée des clients au risque de les voir prendre des distances vis-à-vis du commerce en ligne. Il est difficile de convaincre les clients que ces deux éléments sont pris en compte dans les transactions qu'ils font sur Internet. Ensuite, dans la plupart des cas, les clients n'aiment pas traiter avec un inconnu, sans documents palpables de preuve pour leurs transactions et au moyen de l'argent qui circule par voie électronique. Dans ces conditions, migrer les clients d'un commerce physique vers un commerce virtuel reste très préoccupant.

Une autre limite est le coût toujours élevé de l'accès à Internet. En fait, il faut tenir compte du niveau de vie des populations ; or dans certains pays, la seule solution serait de rendre l'accès à Internet gratuit !

L'aspect légal est une autre limite non technique dans la pratique du commerce électronique ; les gouvernements doivent revoir les lois et règlements devant régir cette nouvelle donne économique. Pour finir, la plupart des clients aiment "toucher du doigt" les biens (habits, parfums, etc.) qu'ils veulent acheter ; sur Internet, cela ne leur est pas possible.

En dépit de ces limites, le commerce électronique connaît son expansion au fil des jours. Aux Etats-Unis [188], le nombre de personnes qui achètent et vendent par la voie électronique est passé de 300 000 en début 1996 à plus de 17 millions en l'an 2001 et autour de 100 millions en 2004. Il est prévu que ce nombre passe à 120 millions en fin 2005. Au Canada, au deuxième trimestre 2002, le nombre moyen d'heures que chaque habitant passait sur Internet était 8,7, contre 12,7 au premier trimestre 2005 [2]. En clair, les expériences accumulées et l'intégration de technologies nouvelles font augmenter le ratio coût-bénéfice du commerce électronique, favorisant ainsi petit à petit son adoption.

[2] Source : The Canadian Inter@ctive Reid Report, Été 2005.

Le commerce électronique se présente sous plusieurs formes, dépendamment des entités en présence et des transactions qui les lient. La prochaine section est consacrée à la présentation des principaux types de commerce électronique pratiqués de nos jours.

1.4 Types de commerce électronique

Il existe actuellement plusieurs champs d'application du commerce électronique, mais nous nous limitons ici à une dizaine d'entre eux.

Business-to-Business ou B2B : Dans ce type, tous les participants sont des compagnies. C'est le cas par exemple des transactions électroniques entre un site Web commercial et une banque avec laquelle il fait affaire. C'est le type le plus pratiqué de nos jours. Les applications les plus connues ici concernent la gestion des fournisseurs (pour améliorer les fonctions d'achats) ; les échanges d'inventaires (pour disposer d'une vue globale de l'offre et de la demande et, si possible, entrer en possession des rapports clients/fournisseurs des autres compagnies de vente) ; la collaboration en temps réel (pour faire face à toute éventualité : inflation, décisions gouvernementales, etc.). `Tradeholding.com` est un exemple de site Web qui pratique du B2B.

Business-to-Consumer ou B2C : Le type B2C fait référence aux ventes faites par les compagnies auprès des clients individuels encore appelés consommateurs. Ce type porte aussi le nom de e-tailing. Les ventes dont il est question ici vont du détail (vêtements, montres, parfums, etc.) aux enchères, en passant par le conseil ou l'assistance (cabinets médicaux, de conseil, d'audits, etc.). `Bell.ca` et `Amazon.com` sont des exemples de sites Web qui pratiquent du B2C.

Les deux types précédents sont les plus importants, voire les plus pratiqués, puisque les échanges entre compagnies et les achats auprès des compagnies disposant des boutiques en ligne sont au cœur même du commerce électronique. Plus concrètement, les transferts électroniques de fonds (EFT) et les échanges de données par les moyens électroniques (EDI) signalés en section 1.2 sont respectivement une forme de B2C et de B2B. Les nouveaux types prennent leurs sources dans ces deux premiers modèles.

Business-to-Business-to-Consumer ou B2B2C : C'est un model dans lequel une compagnie fournit des produits et services à une autre compagnie qui à son tour offre ces services à ses propres clients. Un exemple typique au Canada est la compagnie `B2B2C.ca` qui sous-traite avec `Videotron.ca` pour fournir des services d'Internet à ses clients.

Consumer-to-Business ou C2B : Dans cette catégorie, les individus utilisent l'Internet pour vendre des produits et services à des compagnies ou alors, cherchent des vendeurs et concluent des transactions en ligne avec eux. Un individu détenteur de produits vivriers (légumes, fruits) et qui les vend auprès d'une compagnie faisant des ventes à grande échelle d'une part, et un consultant auprès d'une compagnie d'autre part, sont des exemples de C2B. `Fotolia.com` et `Upload.video.google.com` sont des exemples de sites Web qui pratiquent du C2B.

Consumer-to-Consumer ou C2C : Dans cette catégorie, les consommateurs vendent directement à d'autres consommateurs. Comme exemples, on a : la vente de titres de propriété des résidences, la vente de véhicules, les enchères, la publicité des services personnels, la vente de "connaissances" et autres expertises en ligne, etc. `Ebay.com` est un exemple de site Web C2C. Les sites Web de vente aux enchères (`ubid.com`, `auctions.yahoo.com`, etc.) constituent d'autres exemples.

People-to-People ou P2P : C'est un cas typique du C2C dans lequel les participants échangent par exemple des CDs, des vidéos, des logiciels, etc. Les auteurs de [176] considèrent d'ailleurs le P2P comme une technologie utilisée dans le C2C, le B2B et le B2C et qui permet à des ordinateurs connectés deux à deux (d'où l'expression anglaise *peer-to-peer*) de partager des données. `Kazaa.com` utilise la technologie de pair-à-pair : les différents utilisateurs sont connectés directement entre eux (il y absence totale d'un point central de coordination). L'installation du logiciel *Kazaa* [3] est suffisante pour être relié à d'autres utilisateurs. Par exemple, si Alice et Bob ont tous les deux téléchargé et installé le logiciel Kazaa sur leurs ordinateurs respectifs, alors Bob peut très bien se servir du logiciel Kazaa pour rechercher et télécharger un dossier qui se trouve sur l'ordinateur d'Alice.

Intrabusiness commerce électronique : Ce type fait appel à toutes les activités internes à une compagnie. De telles activités sont faites en Intranet ou à travers des portails et permettent l'échange de biens, de services, d'informations entre plusieurs unités et individus d'une compagnie. En d'autres termes, c'est le commerce électronique pratiqué au sein même d'une compagnie (y compris ses filiales, éloignées ou non, plus ou moins autonomes). Les bibliothèques en ligne (ex. : `librairie.umontreal.ca`) constituent des exemples de ce type.

Business-to-Employees ou B2E : C'est une sous branche de la catégorie précédente avec la particularité que la compagnie délivre les services, les informations ou les produits à ses employés. La compagnie peut par exemple décider de vendre des produits vieillissants à ses employés ou de faire des promotions spéciales à la veille d'une grande fête. Ce type regroupe aussi les services rendus directement aux employés ; par exemple, le site Web `logement.umontreal.ca` met une banque de logements à la disposition des étudiants de l'Université de Montréal.

Government-to-Citizens (ou G2C) and to Others : Dans ce type, une entité gouvernementale achète ou vend des biens, services ou informations aux citoyens, ou à des compagnies. C'est le cas de la vente des timbres fiscaux, des vignettes pour automobiles, etc. Cette catégorie porte aussi le nom de Governement-to-Constituents [26]. Un exemple de site Web qui fait du G2C est `saaq.gouv.qc.ca`.

---

[3]C'est un programme que l'usager installe sur sa machine et qui communique avec le site Web `Kazaa.com`.

Exchange-to-Exchange ou E2E : Les compagnies qui font des échanges ou détiennent des portails électroniques se connectent entre elles pour faire « l'échange de leurs échanges ». L'un des objectifs visés ici peut être la mise à jour des statistiques en vue de l'élaboration de nouvelles stratégies de marketing. `E2esolutions.co.uk` est un exemple de site Web faisant partie de ce type.

Mobile commerce ou M-commerce : Le M-commerce est le commerce électronique pratiqué dans un environnement sans fil (wireless) [26]. Nous pouvons par exemple citer les achats faits à partir de téléphones cellulaires ou d'autres systèmes de transmission de données par satellite. De manière plus spécifique, le commerce électronique pratiqué dans un environnement sans fil inclut les guides d'achats, les itinéraires de voyage, la billeterie, l'actualité, un service personnalisé de films, du divertissement, et biens d'autres services.

Location-based commerce ou L-commerce : Il s'agit d'activités ciblées du m-commerce, à des endroits et temps bien précis. Le L-commerce permet par exemple à un opérateur de déterminer la rue sur laquelle l'on vit et d'envoyer des offres spéciales à partir des magasins environnants. D'ici deux ans [187], en passant à côté d'un McDonald, votre téléphone cellulaire pourrait donner un signal sonore et, en jetant un coup d'oeil sur l'écran, vous pourriez lire : "Spécial deux-pour-un...". Ceci va constituer un autre *Big Brother* [56] que les utilisateurs de téléphones cellulaires n'apprecieraient très probablement pas. Il est intéressant d'émettre le souhait que de tels systèmes ne fonctionneraient qu'avec des ententes préalables avec les utilisateurs.

D'autres types de commerce électronique sont quotidiennement développés. Toutefois, nous nous sommes limité aux types qui précèdent surtout que dans le cadre de cette thèse, nous mettons l'emphase sur le type B2C. Pour l'instant, nous allons nous pencher sur le comportement d'achat d'un client en commerce électronique.

1.5 Modèle CBB (*Customer Buying Behaviour*)

Il existe plusieurs théories et modèles pour décrire le comportement du client dans un processus d'achat [19, 32, 78, 93, 125, etc.]. L'approche de Guttman, Moukas et Maes [91] constitue toutefois un standard dans le contexte du commerce électronique. Les auteurs de [91] définissent dans un premier temps le rôle que les agents logiciels devraient jouer en commerce électronique. Un agent logiciel est un programme informatique avec un but précis, tournant continuellement et semi-autonome. Certaines tâches peuvent leur être déléguées dans un contexte de traitement d'une masse importante d'informations [111], comme c'est le cas du commerce électronique. En effet, la technologie des agents peut être utilisée dans tous les compartiments du commerce électronique : sécurité, mécanismes de paiements, publicité, ontologies, catalogues, intermédiaires, etc. La technologie agent permet essentiellement de sauver du temps car, par exemple, les agents peuvent se charger d'exécuter certaines tâches off-line. Elle peut aussi éviter de passer à côté d'une belle occasion d'affaire, par exemple, une offre d'emploi. Les agents vont donc jouer le rôle important de médiateurs en commerce électronique [21]. En simulant le comportement d'un agent client, Guttman, Moukas et Maes ont défini un comportement d'achat d'un client quelconque en commerce électronique, à travers six étapes.

1. **Identification du besoin :** il s'agit du besoin ou d'une sorte de stimulation pour le client, qui obtient l'information générale ou personnalisée via la publicité sur Internet (y compris, malheureusement, le *spam*). Si le client est inscrit dans une liste de diffusion, `amazon.com` par exemple, il peut recevoir des annonces et publicités par courriel. Dans cette étape, les abonnements en ligne donnent suffisamment d'information sur les intérêts du client. Cette étape porte le nom de `reconnaissance du problem` (*Problem Recognition*) dans le modèle de Engel et Blackwell [78].

2. **Recherche du produit :** cette étape tient lieu de correspondance entre le besoin du client et le produit qui pourrait le satisfaire. Le client critique par lui-même les différents produits qui s'offrent à lui. Il se pose la question "Que dois-je acheter ?"

C'est donc l'étape d'identification du produit.

3. **Recherche du vendeur :** à ce stade, le client recherche un certain vendeur, auprès de qui acheter le produit identifié à l'étape précédente. Généralement, on suppose qu'il a accès à un annuaire Web donnant accès aux vendeurs et aux produits qu'ils vendent. Ici, le client se pose la question "Auprès de qui dois-je acheter ?" Il tiendra particulièrement compte de plusieurs critères : le prix, la disponibilité, le délai de livraison, la garantie, la réputaion du vendeur, etc. Le modèle de Nicosia [125] considère une seule étape, l'`évaluation de la recherche` (*Search Evaluation*) pour parler des deux phases de recherche (produit et vendeur). Engel et Backwell [78] parlent quant à eux de `recherche d'information` (*Information Search*) et d'`évaluation des autres possibilités` (*Evaluation of Alternatives*) pour désigner respectivement la recherche du produit et celle du vendeur.

4. **Négociation :** dans cette étape, le client et le vendeur se servent des stratégies et des fonctions d'utilité pour aboutir à un éventuel accord sur le prix de vente du produit, ainsi que sur les conditions de vente (délai de livraison, garantie, etc.). C'est une étape généralement très interactive, car le client et le vendeur font continuellement des offres et contre-offres dans le but d'aboutir à un consensus.

5. **Paiement et livraison :** cette étape est celle de la matérialisation des transactions, à travers des mécanismes de paiement et la livraison.

6. **Support et évaluation :** c'est la phase qui permet d'avoir un feedback sur le processus d'achat qui lie désormais le client et le vendeur, après les cinq précédentes étapes. Il s'agit en d'autres termes de l'étape du service après vente et de la maintenance, très connu de nos jours sous l'appellation "soutien technique" ou "service à la clientèle". On parle aussi souvent de "e-service" (ex. : chez `Bell.ca`).

Dans le cadre de cette thèse, quand nous parlons du modèle CBB, il s'agit de modèle de Guttman, Moukas et Maes que nous venons de présenter. C'est le modèle le plus utilisé en commerce électronique, aussi sommes-nous partis de lui pour élaborer un nouveau modèle qui préserve la vie privée du client (chapitre 4). En effet, c'est au travers du modèle CBB que le comportement du client est capturé et que la plupart de ses données personnelles sont collectées. Comme nous le verrons au chapitre 3, cette capture et cette collecte constituent l'essence même de la violation de vie privée observée sur Internet en général et en commerce électronique en particulier.

Les phases de recherche du produit et du vendeur sont souvent complexes à cause de la grande masse d'informations disponibles sur Internet. Dès lors, il est important de trouver des outils qui peuvent aider le vendeur à proposer efficacement des produits à ses clients en fonction de leurs besoins et intérêts. C'est dans cette optique que les systèmes de recommandation ont vu le jour. Nous allons en parler dans la prochaine section.

1.6 Systèmes de recommandation

Depuis l'avènement d'Internet, la quantité d'informations disponibles en ligne devient de plus en plus importante. Un client à la recherche d'un produit peut donc facilement être submergé de données et/ou avoir de la difficulté à faire des choix. Pour pallier à cela, les systèmes de recommandation ont vu le jour. Leur but principal est de faciliter la tâche de recherche de produits qui pourraient satisfaire le client selon ses besoins et intérêts. Cette section fait un survol des différents systèmes de recommandation.

1.6.1 Généralités

Un Système de Recommandation (SR) en commerce électronique est un outil de recherche d'informations qui permet à un client de choisir auprès du vendeur les produits qui répondent le mieux à son besoin. En particulier, un SR va guider le client dans ses choix en se servant éventuellement de son profil en termes de données démographiques et d'habitudes d'achats et de navigation.

Un SR se sert toujours d'une technique de filtrage qui prend en entrée une requête du client (par exemple la description de son besoin, combiné à ses précédentes interactions avec le SR) et qui propose en sortie, à l'intention du client une liste de produits, encore appelée "produits recommandés" ou, tout simplement, "recommandations". Les techniques de filtrage peuvent être classées en fonction de plussieurs critères [44] : les données sur lesquelles la technique travaille, l'utilisation qui en est faite, etc. Dans tous les cas, le souci d'affiner les recommandations fait du SR un demandeur d'informations du client, au fur et à mesure de ses passages (dans le SR). En particulier, le SR peut s'intéresser au feedback que le client a des recommandations qui lui ont été faites. Le feedback quant à lui peut être implicite ou explicite.

Le *feedback implicite* est une option qui permet au SR de recueillir de l'information sur le client à son insu, à travers son comportement vis-à-vis du système. Dans ce cas, l'utilisation du SR est transparente pour le client : il n'a pas besoin d'évaluer par lui-même les recommandations reçues. Le SR travaille uniquement avec les données relatives aux achats effectués par le client, aux produits consultés, à la durée de consultation, aux données issues des cookies, log files, spywares, web bugs (section 3.6), etc. Il faut toutefois signaler que les données issues du feedback implicite sont moins pertinentes [34, 124] que celles provenant d'un feedback explicite.

Le *feedback explicite* sollicite du client une évaluation des produits qui lui sont recommandés. Cette évaluation peut être textuelle, s'il est demandé aux clients de préciser leurs opinions par écrit sur les produits recommandés. Pour plus de rigueur, le feedback explicite est souvent recueilli à travers des évaluations sur une échelle numérique binaire (0 et 1), de 1 à 5, de 1 à 10, etc. La principale difficulté avec le feedback explicite est que les clients peuvent faire des évaluations par "plaisanterie", c'est-à-dire qu'il est difficile de mesurer le degré de sérieux accordé par les clients aux évaluations qu'ils font. En clair, un feedback explicite exige du temps et un certain effort cognitif de la part du client. La bonne nouvelle est toutefois qu'il permet d'améliorer la qualité des recommandations, surtout si la majorité des clients font des évaluations sérieuses.

Dans tous les cas, le profil du client donne lieu à la personnalisation du contenu des pages Web que le vendeur met à sa disposition d'une manière générale.

1.6.2 Profil de client

Le profil de client a quatre composantes essentielles :

- L'identité : ce sont les données nominatives du client.
- Les données démographiques : il s'agit de l'information personnelle sur le client.

L'âge, le sexe, l'adresse, le revenu annuel, le niveau d'éducation sont quelques exemples de données démographiques.

- Les habitudes d'achats : ici, l'intérêt porte sur ce que le client a acheté par le passé ; les lieux où les achats ont été effectués ; les montants dépensés ; les modes de paiement utilisés ; la sensibilité du client à la marque des produits ; etc.
- Les habitudes de navigation : elles portent sur le comportement du client face à une page Web. Il est par exemple possible d'enregistrer toutes les pages Web visitées par le client pour en faire une analyse. Un tel enregistrement porte le nom de *chaîne des clics* ou, en anglais, *clickstream*.

Pour le cas précis des SR, les *techniques de filtrage* vont se servir du profil du client pour ajuster les recommandations, dépendamment du contenu d'un tel profil.

1.6.3 Quelques techniques de filtrage

Dans un récent papier d'Adomavicius et Tuzhilin [4], les auteurs considèrent trois types de SR prenant appui sur le filtrage par contenu, le filtrage collaboratif et l'approche hybrique des deux premiers. Trois années plus tôt, Burke [44] avait déjà distingué cinq principales techniques de filtrage utilisées en commerce électronique : le filtrage par contenu, le filtrage collaboratif, le filtrage démographique, le filtrage basé sur l'utilité et le filtrage basé sur les connaissances. Ci-dessous, nous décrivons sommairement l'approche de Burke qui, à notre avis, se veut plus englobante, à travers les cinq techniques précédentes. Nous parlerons aussi des différentes associations qu'on peut faire de ces techniques pour produire des *hybrides* [44].

1.6.3.1 Le filtrage collaboratif

Le filtrage collaboratif ou CF (*Collaborative Filtering*) compare les clients, en se basant sur les achats qu'ils ont effectués par le passé et génère des prédictions qui tiennent compte de la similarité entre clients. Dans un SR utilisant la technique CF, le processus de recommandation présente trois grandes étapes :

1. Chaque utilisateur donne un *vote* ou une *évaluation* sur différents produits.

2. Le SR recherche les clients les plus similaires en basant la fonction de similarité sur les votes. Si deux clients ont donné les mêmes votes ou des votes proches sur les mêmes produits, alors ces deux clients donneront *a priori* des votes similaires sur d'autres produits.
3. De nouvelles recommandations sont générées en fonction des clients les plus similaires.

Il existe deux classes principales de filtrage collaboratif [41] : les algorithmes basés sur la mémoire (*memory-based*) et ceux basés sur les modèles (*model-based*).

Les algorithmes basés sur la mémoire utilisent tous les profils utilisateurs afin de générer de nouvelles prédictions. En général, on distingue deux étapes principales dans l'exécution de tels algorithmes. D'une part, ils recherchent les clients ayant évalué les mêmes produits de façon similaire au client courant, c'est-à-dire le client pour qui on souhaite faire des recommandations. D'autre part, ces algorithmes doivent calculer, en fonction des votes des clients similaires, des prédictions sur les produits que le client courant n'a pas encore évalués. Plusieurs techniques permettent de calculer la similarité entre deux clients. La plus populaire des techniques est la *corrélation de Pearson* [140].

La corrélation de Pearson est une technique qui a été utilisée dans les systèmes de recommandation pour la première fois dans le cadre du projet GroupLens en 1994 [140]. La corrélation entre deux clients, c et u, est définie par l'équation 1.1 :

$$corr(c, u) = \frac{\sum_{j \in J}(v_{c,j} - \overline{v_c})(v_{u,j} - \overline{v_u})}{\sqrt{\sum_{j \in J}(v_{c,j} - \overline{v_c})^2 \sum_{j \in J}(v_{u,j} - \overline{v_u})^2}} \tag{1.1}$$

Dans l'équation 1.1, J est l'ensemble des produits évalués à la fois par c et u, $v_{i,j}$ est le vote du client i sur le produit j et $\overline{v_i}$ représente la moyenne des votes du client i sur les produits pris dans J, pour $i \in \{c, u\}$.

La prédiction de vote, $P_{c,j}$, pour un client c sur un produit j, peut alors être calculée en utilisant, par exemple, l'algorithme des k meilleurs voisins ou kNN (*k-Nearest Neighbours-based algorithm*) [67,68]. Le calcul de prédictions avec kNN utilise l'équation 1.2 :

$$P_{c,j} = \overline{v_c} + \frac{\sum_{u \in U} corr(c,u)(v_{u,j} - \overline{v_u})}{\sum_{u \in U} |corr(c,u)|}, \qquad (1.2)$$

où U est l'ensemble de tous les clients qui ont donné un vote sur le produit j. (En particulier, le client c, pour qui les prédictions sont calculées, ne fait pas partie de l'ensemble U). En fin de compte, les valeurs $P_{c,j}$ sont triées et les produits, j, faisant partie des k valeurs (de $P_{c,j}$) les plus élevées sont recommandés au client c.

Un algorithme à base de modèle utilise les profils des clients pour estimer ou apprendre un modèle, qui permettra ensuite de faire des prédictions pour un client donné. En fait, un algorithme à base de modèle permet de mettre sur pied un modèle du client, puis utilise une approche probabiliste pour faire des prédictions : il génère la probabilité qu'un client c ait un vote $v_{c,j}$ sur un produit j, sachant ses votes précédents. Il existe plusieurs techniques de construction de modèles, parmi lesquels se retrouvent le modèle de réseaux bayesiens [41, etc.] et le modèle de clusters [24, 41, etc.].

Dans cette thèse, nous avons essentiellement travaillé avec les algorithmes basés sur la mémoire.

Le CF reste de nos jours la technique de filtrage la plus utilisée dans les SR. Toutefois, cette technique présente plusieurs défauts. Notre intérêt a porté sur quatre défauts à savoir le *cold start*, la *sparsity*, le *gray sheep* et le *shilling*, que nous présentons ci-dessous. Pour une liste plus complète des difficultés que connaissent les SR, le lecteur est invité à consulter [4, 44] par exemple.

- ***Cold Start*** ou **Démarrage à froid** : il est difficile de recommander un nouveau produit qui n'a pas encore reçu d'évaluations (votes) de la part des clients. Ce problème se pose avec acuité dans un contexte de commerce électronique où les nouveaux produits font constamment leur entrée dans les catalogues des vendeurs. Il faut d'ailleurs noter que si le nouveau produit a reçu un petit nombre d'évaluations, le problème va persister, car plus il y a d'évaluations sur un produit, meilleures sont les corrélations entre les clients (évaluateurs) [167]. Diverses solutions sont proposées pour contourner ce problème, dépendamment du domaine (films, musique, etc.) de prédilection du SR. Par exemple, Melville *et al.* [116]

proposent l'obtention des prédictions basées sur le contenu à partir du voisinage du client.

Le démarrage à froid fait aussi allusion à la difficulté de catégoriser les nouveaux clients. Ces derniers n'ont pas encore fait d'évaluations de produits, il n'existe donc pas de corrélations entre eux et les clients du système. Dans ce cas, le SR doit motiver le nouveau client à évaluer plusieurs produits. Par exemple, les auteurs de [15,73] proposent aux usagers de leurs systèmes d'évaluer quelques films qu'ils ont déjà vus pour que le système soit en mesure de leur fournir de meilleures recommandations.

- ***Sparsity*** : lorsque le catalogue du vendeur est très riche et varié en produits, il se peut que peu de clients aient voté sur les mêmes produits. En d'autres termes les votes des clients sont épars et il devient difficile d'établir des corrélations entre clients. Une solution à ce problème consiste à créer des grappes (clusters) de clients, qui sont en fait des stéréotypes dans lesquels il faut classer un client donné selon son profil et ainsi réduire le nombre de ses voisins. Ceci permet de renforcer les corrélations entre clients [167, 171].
- ***Gray sheep*** : le CF tend à recommander les produits les plus populaires, c'est-à-dire les produits pour lesquels le SR dispose de plus de bonnes évaluations. Ainsi, si un client a des goûts particuliers (différents de ceux de la masse des clients), il devient difficile de générer de bonnes recommandations pour lui, car il est difficile d'établir son voisinage [62].
- ***Shilling*** : il se peut que des personnes malveillantes créent plusieurs profils pour se connecter au SR afin de l'influencer à travers les votes des différents profils. Par exemple, un fabriquant pourrait chercher à ce que ses produits soient plus recommandés que ceux de ses concurrents. En créant plusieurs profils de clients, il va évaluer positivement ses produits et négativement ceux de la concurrence, augmentant ainsi de manière déloyale les chances de voir ses produits recommandés à tous les coups. Pour plus de détails sur ce problème et les solutions proposées, il serait intéressant de consulter [59, 106, 128].

En somme, les SR qui utilisent le CF sont efficaces car ils produisent de bonnes recommandations pour le client. En particulier, plus le nombre d'évaluations est grand sur un grand nombre de produits, meilleures sont les recommandations. Toutefois, les problèmes présentés ci-dessus (cold start, sparsity, gray sheep, shilling) constituent une limite difficilement surmontable par le CF, sans oublier le fait que des compagnies peuvent négocier avec des systèmes de recommandation pour que leurs produits soient les plus recommandés.

1.6.3.2 Le filtrage basé sur le contenu

Le filtrage basé sur le contenu ou CN (*Content-Based Filtering*) compare les produits en se basant sur leur contenu, c'est-à-dire sur les attributs ou caractéristiques des produits. Par exemple, pour un film, le contenu peut être lié au type (action, aventure, etc.), l'acteur principal (Bruce Willis par exemple), l'année de parution, etc. Il peut aussi s'agir d'un contenu textuel (cas d'un livre avec les attributs titre, description, table des matières, etc.). Dans ce dernier cas, différentes techniques (par exemple, le modèle vectoriel de Salton [149, 151]) permettent de calculer des *similarités* entre deux documents textuels donnés.

Le CN fournit des recommandations en utilisant les caractéristiques des produits et l'intérêt que les clients portent sur elles. Ainsi, un SR qui est basé sur cette technique va suggérer au client des produits ayant des caractéristiques similaires à celles des produits qu'il a évalués [4] pour signifier qu'il les a aimés ou achetés dans le passé.

Le CN est aussi connu sous le nom de "corrélation *produit-à-produit*" (*product-to-product correlation*) [155]. C'est pourquoi on peut se servir d'un algorithme qui calcule la similarité produit-à-produit [153] dans un SR utilisant le CN. Pour cela, les produits peuvent être considérés comme des vecteurs dans un espace multidimensionnel et la similarité entre deux produits i et j est calculée suivant l'équation 1.3 :

$$Sim(i,j) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\| \cdot \|\vec{j}\|} \tag{1.3}$$

[4] Les évaluations se font généralement par le biais des votes sur une échelle de 1 à 5, de 1 à 10, etc.

L'équation 1.4 est par la suite utilisée pour calculer la prédiction pour un client c et un produit j :

$$P_{c,j} = \frac{\sum_{i \in S} v_{c,i}\, Sim(i,j)}{\sum_{i \in S} |Sim(i,j)|} \tag{1.4}$$

où S est l'ensemble des tous les produits similaires à j et $v_{c,i}$ est définie comme à l'équation 1.1.

Le CN présente un double avantage : d'une part, il n'a plus besoin de tenir compte de l'historique des achats effectués par le client et, d'autre part, il est facile à implanter. Par contre, le CN fournit des recommmandations "statiques" puisque l'ensemble des caractéristiques ne change presque pas. De plus, si le contenu est vague (cas des livres), le CN peut être très inefficace.

1.6.3.3 Le filtrage démographique

Les SR basés sur le filtrage démographique ou DF (*Demographic Filtering*) visent à classer les clients par catégorie, dépendamment de leurs données démographiques. Ils recommandent alors les produits à travers cette catégorisation. Plus précisément, les données démographiques servent à identifier les types de clients qui aiment les produits similaires. L'élément principal du filtrage démographique est qu'il crée des catégories de clients ayant des caractéristiques démographiques similaires et analyse les comportements d'achat ou les préférences globales des clients faisant partie ce ces catégories. Pour faire des recommandations de produits à un nouveau client, le SR doit d'abord trouver la catégorie à laquelle il appartient, puis appliquer les préférences d'achat des clients qui s'y trouvent. Comme exemple de SR utilisant le DF, on a *LifeStyle Finder* [102] de Krulwich. LifeStyle Finder utilise les informations démographiques des usagers pour les classer dans l'un des 62 stéréotypes définis dans le cadre d'une recherche de marketing. À partir de ces classes d'usagers et des caractéristiques connues pour chacune d'elles, LifeStyle Finder est capable de produire des recommandations.

L'avantage d'utiliser le DF est le fait de ne pas imposer aux clients d'évaluer un certain nombre de produits avant de pouvoir générer des recommandations. En travaillant au niveau des agrégats, le DF est très performant et peut ainsi servir de première étape à franchir dans un processus de filtrage collaboratif (section 1.6.3.1) pour former un voisinage initial du client. Par contre, le principal inconvénient est que la similarité démographique ne met pas en relief les différences entre les clients, ce qui conduit à des recommandations très générales. De plus, le profil du client étant fixe, ses changements de goûts ne sauraient être pris en compte [120].

Le clustering [94] fait partie des techniques utilisées pour faire de la catégorisation. Il a fait l'objet de beaucoup de recherche, en particulier dans les domaines tels que les statistiques, l'apprentissage machine, le forage des données (*data mining*), etc. Le clustering vise à former des agrégats d'objets, de telle sorte que les objets dans un agrégat donné sont similaires entre eux, mais différents de ceux d'un autre agrégat. Un algorithme de clustering a ainsi besoin d'une métrique pour calculer la distance entre deux objets et relever jusqu'à quel point ils sont similaires ou différents. Chaque agrégat dispose généralement d'un objet de référence connu sous le nom de *centroïde* ou de *méthoïde* pour le représenter. Ainsi, quand on est en présence d'un nouvel objet à catégoriser, il suffit de calculer sa distance par rapport à l'objet de référence de chaque agrégat, puis de choisir l'agrégat correspondant à la distance minimum.

1.6.3.4 Le filtrage à base d'utilité

Le filtrage basé sur l'utilité ou UF (*Utility-Based Filtering*) exige qu'une fonction d'utilité soit attribuée à chaque client. Un SR basé sur cette technique calcule alors l'utilité de chaque produit pour faire des recommandations au client. Comme le filtrage basé sur le contenu, les fonctions d'utilité utilisent des informations sur les produits. Toutefois, ces informations ne se limitent pas seulement au produit en soi ; elles peuvent porter sur d'autres facteurs tels que le prix, le délai de livraison etc. En d'autres termes, le système tient largement compte des contraintes exprimées par le client afin de lui procurer des recommandations qui répondent à son besoin immédiat. Le système Tête-à-Tête [91] développé par Guttman est un exemple de système qui utilise le UF.

1.6.3.5 Le filtrage à base des connaissances

Le filtrage basé sur les connaissances ou KF (*Knowledge-Based Filtering*) considère les besoins et préférences des clients pour en déduire des recommandations. Les connaissances dont il est question ici peuvent être "fonctionnelles", c'est-à-dire que le SR a des connaissances sur comment un produit particulier peut correspondre aux besoins d'un client. Par exemple, le système "Entrée" de Burke [44] utilise ses connaissances sur la cuisine pour déduire des similarités entre des restaurants.

D'une manière générale, le KF transforme les achats effectués dans le passé par les clients en des règles d'associations du genre, "les clients qui ont achetés le produit P_1 ont aussi acheté le produit P_2". Le KF utilise alors ces règles d'association pour faire des recommandations.

Les SR basés sur le KF présentent un certain nombre d'avantages : ils sont faciles à implanter, rapides ; il ne requièrent pas d'espace de stockage additionnel et ne dépendent pas d'un client particulier. Comme inconvénients, il peut être difficile au KF de s'adapter à un changement brusque de la connaissance, par exemple un changement de nom d'un produit.

Il faut noter qu'il y a une différence fondamentale entre un SR basé sur le CN et un SR basé sur le KF. Pour le premier, les produits sont comparés en se basant sur leurs contenus, tandis que pour le second, on se sert des achats faits dans le passé comme métrique de comparaison des produits.

1.6.4 Les techniques hybrides

Les techniques hybrides résultent en la combinaison d'au moins deux techniques de filtrage dans le but d'améliorer la performance du SR, en essayant en particulier de contourner les problèmes posés par une seule technique [44]. Voici quelques types d'hybridation présentés dans [44, 101] : pondération, commutation, mixte, combinaison de caractéristiques, cascade, augmentation de caractéristiques et méta-niveau. Nous les définissons ci-dessous.

1.6.4.1 Pondération (*Weighted*)

Les résultats produits (votes ou scores) par différentes techniques de recommandation sont combinées de manière à produire une seule recommandation [62, 129].

1.6.4.2 Commutation (*Switching*)

Cette technique permet de choisir une technique de recommandation entre plusieurs selon certains critères. La détermination de la technique à utiliser dépend de la situation. Le système doit donc définir les critères de commutation (dans quels cas utiliser une autre technique). Cependant, cela permet au système de connaître les forces et faiblesses des techniques de recommandations qui le constituent. Les systèmes présentés dans [72, 174] sont des exemples d'hybrides par commutation.

1.6.4.3 Technique mixte (*Mixed*)

La technique mixte permet de donner à l'utilisateur des recommandations provenant de plusieurs techniques simultanément. Un exemple de technique mixte est l'utilisation des techniques de filtrage collaboratif et de filtrage par contenu. Cela permet d'éviter un des problèmes du filtrage collaboratif : le "démarrage à froid" du nouvel item. En effet, la technique basée sur le contenu permet d'obtenir des recommandations des nouveaux objets en se basant sur leurs descriptions. Le système PTV [166] qui permet d'établir un horaire télévisé personnalisé pour un client donné est un exemple de SR de type hybride mixte.

1.6.4.4 Combinaison de caractéristiques (*Features combination*)

Dans cette technique, les caractéristiques des informations fournies par les différentes méthodes de recommandation sont combinées pour permettre l'utilisation d'une seule technique sur l'ensemble de ces données. Par exemple, le filtrage collaboratif utilise principalement les votes des utilisateurs comme sources de données, elles peuvent être ajoutées aux informations utilisées par le filtrage par contenu pour produire des recommandations. Le système décrit dans [24] illustre la technique hybride par combinaison de caractéristiques.

1.6.4.5 Cascade (*Cascade*)

Cette technique se déroule en deux étapes : une première technique est utilisée pour produire un ensemble de candidats potentiels, et la deuxième technique sert à raffiner les recommandations faites. L'avantage de cette méthode hybride est que la deuxième technique ne servira que si les recommandations produites par la première nécessitent une discrimination supplémentaire (dans le cas où les recommandations sont trop nombreuses par exemple). Par contre, si la première technique donne trop peu de recommandations, ou si celles-ci sont déjà ordonnées pour permettre une sélection rapide, alors la deuxième technique ne sera pas utilisée. Le système Fab [23] qui fait la recommandation de documents textuels sur Internet, est un exemple d'hybridation en cascade.

1.6.4.6 Augmentation de caractéristiques (*Feature augmentation*)

Dans ce type, une première technique de filtrage est utilisée pour produire un rang ou une évaluation pour chaque produit et la deuxième technique se sert de cette information pour opérer. Cet hybride ressemble donc beaucoup à la cascade, la différence entre les deux réside en ce que le résultat de la première technique est prise en compte dans le calcul effectué par la deuxième technique. Le système Libra [121], utilisé pour la recommandation de livres, est un exemple d'hybride de type augmentation de caractéristiques.

1.6.4.7 Méta-niveau (*Meta-level*)

Ici, le modèle complet issu d'une technique de filtrage qui est utilisé par une autre technique. Le système Fab [23] a été le premier système hybride méta-niveau. D'autres systèmes ont vu le jour par la suite [129].

En plus de la recherche de produits, juste avant la phase de paiement, le client et le vendeur négocient généralement le prix de vente final du produit.

1.7 Négociation

La négociation [61, 95–97, 110] en commerce électronique est un processus dans lequel les participants (vendeurs et clients) sont à la recherche d'un accord sur le prix de vente final du produit, ainsi que les conditions de vente (garantie, remboursement, échange, etc.). Généralement, elle fonctionne en termes d'offres et contre-offres. Les enchères électroniques constituent un exemple de tâche qui nécessite la négociation. La négociation comprend trois principales composantes :

- Le *protocole* : c'est l'ensemble des règles qui régissent la négociation, en l'occurrence, les types d'offres acceptables. Par exemple, dépendamment du type d'enchères, une règle peut consister à ne pas diminuer/augmenter les offres faites.
- La *stratégie* : elle est propre à chaque participant. C'est une spécification de ce qui doit être fait pour chaque situation qui pourrait arriver pendant le processus de négociation. La stratégie constitue en quelque sorte une *boîte à décisions* pour le négociant. Les décisions vont consister à produire ou à accepter des offres/contre-offres. Elles utilisent des *fonctions d'utilité*, qui ont justement pour rôle de calculer l'utilité d'une offre donnée. En d'autres termes, la décision vis-à-vis d'une offre est prise sur la base de sa valeur d'utilité.
- Les *objects* : ce sont les ressources sur lesquelles la négociation porte. En commerce électronique, ces objets sont par exemples des produits. De manière générale, un objet est caractérisé par une liste d'attributs (prix, garantie, etc.), de sorte que les offres vont consister en des valeurs (100$, 24 mois, etc.) affectées à ces attributs.

Il existe plusieurs types de négociation parmi lesquels :

- *Absence de négociation* : le vendeur offre ses produits en se servant d'une liste de prix fixes, avec la formule de vente : "vous prenez ou vous laissez" [103].
- *Enchères* : les offres sont faites sous forme mises. Les enchères commencent toujours avec des mises initiales qui vont progressivement augmenter (enchères anglaises) ou diminuer (enchères hollandaises) par la suite. `Ebay.com` et `ubid.com` sont des exemples de sites Web qui font des enchères.

- *Bilatérale* : la négociation se passe entre deux participants (généralement entre un vendeur et un client) qui font des offres/contre-offres jusqu'à ce qu'un consensus soit atteint, ou jusqu'à ce que le processus soit annulé par au moins l'un des participants. Dans cette thèse, nous proposons une *négociation masquée* de type bilatéral en section 5.4.

D'autres types de négociation ainsi qu'une analyse plus poussée des types précédents peuvent être consultés dans [30].

1.8 Conclusion

Dans ce chapitre, nous avons fait un survol du commerce électronique. Nous sommes loin d'avoir couvert un pourcentage significatif de ce vaste champ de recherche. Nous avons juste voulu présenter les rubriques qui ont un rapport plus ou moins direct avec notre travail. Si la section historique nous as ramenés à l'origine du e-commmerce, ses fondements présentés en section 1.3 nous ont enrichis d'un vocabulaire qui lui est propre. Toujours en section 1.3, nous avons mis en exergue les différents acteurs et mécanismes, ainsi que les disciplines, pressions, bénéfices et limites du commerce électronique. Par la suite, nous avons décrit en section 1.4 les divers types de commerce électronique pratiqués de nos jours avant de déboucher, en section 1.5, sur le modèle CBB dont le but est de capturer le comportement du client quand il effectue des transactions d'achats en ligne. La section 1.6 a porté sur les systèmes de recommandation qui, dans la plupart des cas, utilisent le profil du client pour lui recommander les produits pouvant satisfaire ses desiderata. D'une manière générale, le profil du client [173] va rélever :

- des données statiques : qui est le client ? que possède-t-il ? etc.
- des données dynamiques : les transactions conduites en ligne par le client, son mode de vie, sa localisation, etc.
- des données dérivées : habitudes d'achats, de navigation, habitudes sociales, etc.

Nous reviendrons sur le profil du client au chapitre 3. Pour finir, en section 1.7, nous avons donné un aperçu de la négociation en commerce électronique.

Au chapitre 4, nous nous servons des protocoles et primitives cryptographiques pour introduire notre modèle d'achats qui préserve la vie privée des clients, ainsi que les données sensibles du vendeur, en commerce électronique. Il est donc important d'introduire tous ces concepts. Ceci est fait au chapitre 2, consacré à la cryptographie.

CHAPITRE 2

CRYPTOGRAPHIE

Une des caractéristiques les plus singulières dans l'art de chiffrer un message est la forte conviction qu'a chaque personne d'être capable de construire un système de chiffrement que personne d'autre ne pourra jamais déchiffrer

C. Babbage

2.1 Introduction

Pour lutter contre la violation de vie privée des clients et protéger les données sensibles du vendeur dans le contexte du commerce électronique, nous avons choisi une approche cryptographique (voir chapitres 4 et 5). Dans le présent chapitre, nous allons faire le tour des différents systèmes crytographiques et de leurs applications.

Le mot cryptographie vient des mots grecs *kruptos* et *graphein*, qui signifient respectivement "caché" et "écrire". Ainsi, la cryptographie est l'ensemble des procédures qui permettent de transformer un message intelligible ou texte clair, en un texte chiffré incompréhensible ou *cryptogramme*, à travers des algorithmes préalablement convenus entre l'émetteur et le destinataire du message. Ces algorithmes sont au nombre de trois : d'abord, l'algorithme de *chiffrement*, qui permet de transformer le texte en clair en un texte chiffré et, d'autre part, l'algorithme de *déchiffrement*, qui permet de retrouver le texte en clair à partir du texte chiffré. Par ailleurs, le chiffrement et le déchiffrement utilisent une information additionnelle appelée *clé*, produite par le biais de l'algorithme de *génération de clé*. Les trois algorithmes forment ce qu'on appelle *système cryptographique* ou *cryptosystème*. Ce dernier utilise généralement trois espaces de textes :

- un espace des textes en clair (messages) : ce sont des textes écrits dans un langage accessible à un certain groupe d'individus (exemples de langages : anglais, français, braille, etc.).

- un espace des textes chiffrés (cryptogrammes) : ce sont des textes écrits dans un langage secret, accessible seulement à un petit nombre d'individus aidés par les clés dont ils disposent.
- un espace des clés (clés secrètes, privées et publiques)[1] : c'est un ensemble à partir duquel les clés sont choisies.

De manière plus précise, dans la phase de chiffrement, un cryptosystème combine [39, 169] le texte en clair et une information additionnelle appelée *clé* pour produire le texte chiffré. En retour, le texte chiffré est aussi combiné à une clé pour retrouver le texte en clair.

Un *cryptographe* est, entre autres[2], un spécialiste du chiffrement/déchiffrement ; il crée et améliore des codes mathématiques utiles au développement des algorithmes de chiffrement/déchiffrement. Tout à son opposé, se retrouve le *cryptanalyste*, chargé de briser ces codes et d'accéder au contenu des messages confidentiels. L'ensemble des procédés permettant de briser les codes des cryptographes et/ou de retrouver des clés secrètes de manière illégitime s'appelle la *cryptanalyse*. La cryptographie forme avec la cryptanalyse un domaine de la science appelé *cryptologie*.

Pour qu'un cryptosystème soit sécuritaire, il doit être très difficile de déduire le texte clair à partir d'un texte chiffré, sans disposer de la clé secrète. En pratique, cette demande se limite au fait que le système est difficile à casser. La cryptographie a pour but d'assurer la sécurité des communications et des données stockées en présence d'un *adversaire*[3]. Elle ne permet pas de résoudre tous les problèmes liés à la sécurité informatique ou électronique. Par exemple elle ne peut rien contre les attaques par les virus informatiques qui profitent parfois de la confiance exagérée des utilisateurs dans les messages ou logiciels qu'ils reçoivent. Mais, la cryptographie constitue une base à partir de laquelle des protocoles de sécurité peuvent être construits. Plus précisément, elle participe à la sécurité des systèmes informatiques en proposant des primitives permettant

[1] Voir, section 2.4.

[2] Le cryptographe peut être spécialiste des fonctions de hachage ou signatures numériques. Ces deux notions sont introduites dans la suite de ce chapitre (Section 2.4.7).

[3] Espion, malfaiteur, etc.

d'atteindre des objectifs telles que[4] la *confidentialité*, l'*intégrité*, l'*authentification*, la *signature numérique*, la *non-répudiation* ou l'*identification*, etc. La révolution de l'Internet et l'utilisation de plus en plus massive d'informations sous forme numérique facilitent les communications et rendent de ce fait plus fragiles les informations que l'on détient. En effet, les réseaux ouverts tels que l'Internet créent des failles de sécurité et il est plus aisé à un adversaire d'accéder aux informations. Le remplacement de l'homme par la machine rend les relations beaucoup plus anonymes alors qu'en même temps, l'accès aux données demande des moyens d'authentification forts. La révolution numérique des communications et de l'information a ainsi ouvert de nombreux champs d'investigation à la cryptographie, de sorte que celle-ci a envahi notre vie quotidienne (transaction bancaire, Internet, téléphone mobile, etc.).

Les cryptosystèmes existent depuis environ 4000 ans, c'est-à-dire depuis au moins les dérivées hiéroglyphes utilisées en Egypte, (1900 av. JC). Dans ce chapitre, nous allons présenter la cryptographie et ses applications. En particulier, nous mettrons en relief l'utilisation de la cryptographie en commerce électronique.

Ce chapitre est organisé comme suit : la section 2.3 est consacrée aux attaques, aux risques et à la protection des données d'une manière générale, mais spécifiquement dans un contexte de commerce électronique. En section 2.2, nous présentons quelques primitives cryptographiques. Nous pousuivons avec une description de la notion de crytographie en section 2.4. Par la suite, en section 2.5, nous mettons en exergue l'utilisation de la cryptographie en commerce électronique, avant de conclure le chapitre en section 2.6.

2.2 Primitives cryptographiques

De nombreuses primitives mathématiques jouent un rôle de premier plan dans la conception des solutions cryptographiques. Il s'agit notamment des fonctions à sens unique, des fonctions de hachage, des générateurs aléatoires et pseudo-aléatoires, etc. Cette section est consacrée à la description de quelques unes de ces primitives.

[4] Voir, section 2.3.

2.2.1 Fonctions à sens unique

Les fonctions à sens unique sont des fonctions faciles à calculer dans un sens mais difficiles à calculer dans l'autre sens. Plus concrètement, si f est une fonction à sens unique, alors étant donné x, il est facile de calculer $y = f(x)$ mais, étant donné y, il est difficile de trouver x tel que $f(x) = y$. On pense que de telles fonctions existent mais personne ne sait le prouver. Si f est une fonction à sens unique, alors, pour un x donné, la valeur $f(x)$ porte le nom de *digest*.

Voici deux exemples de fonctions conjecturées comme fonctions à sens unique.

- Étant donné deux grands nombres premiers, p et q, on pense que la fonction $f(p, q) = pq$, est à sens unique car, s'il est facile de calculer $n = pq$, l'opération inverse consistant à trouver p et q à partir de n est plutôt difficile à effectuer.
- Étant donné un nombre *composé*, c'est-à-dire $n = p_1 \times p_2 \times \cdots \times p_k$ où $p_1, p_2, \cdots, p_k$ sont de grands nombres premiers inconnus, on pense que la fonction $g(x) = x^2 \bmod n$ qui calcule des carrées modulo n est une fonction à sens unique.

2.2.2 Fonctions de hachage

Une fonction h est dite de hachage si elle est facile à calculer et si elle fait correspondre à un ensemble X composé de chaînes de bits de longueur finie mais arbitraire, un ensemble Y, de taille inférieure à celle de X, composé de chaînes de bits de longueur finie et fixée. Comme dans le cas d'une fonction à sens unique, le résultat d'une fonction de hachage porte aussi le nom de "digest".

Une fonction de hachage est dite à clé si une clé secrète intervient dans son calcul ; sinon, elle est sans clé. Les fonctions de hachage ont de nombreuses applications informatiques parmi lesquelles l'archivage structuré qui facilite la recherche ou l'extraction de données. Sur le plan de la sécurité, on distingue deux classes principales de fonctions de hachage.

D'une part, les codes détecteurs d'altérations ou *MDC* (*Manipulation Detection Codes*) or encore *MIC* (*Message Integrity Codes*). Ces codes sont "sans clé" et peuvent être utilisés dans un service d'intégrité. Parmi les *MDC* connus, on peut citer [169] :

le MD4 et le MD5 (digest de 128 bits pour l'un et l'autre), développés par Ronald Rivest [143, 145] respectivement en 1990 et 1992 ; le SHA-256 (digest de 256 bits), développé par le NIST (*National Institut of Standards and Technology*) en 1995. Les écritures abrégées MD4 et SHA signifient respectivement *Message Digest [version] 4* et *Secure Hash Algorithm.*

D'autre part, on a les codes d'authentification de message ou MAC (*Message Authentication Codes*). Ils sont "à clé" et permettent d'authentifier la source d'un message et de s'assurer de son intégrité (en autant de connaître la clé) sans utiliser de mécanismes additionnels tel que le chiffrement. Le MD5-MAC est un exemple de code MAC. D'autres exemples peuvent être consultés dans [118, 169].

La principale application des fonctions de hachage est d'assurer l'intégrité de l'information. Pour cela, on peut soit se servir uniquement du MAC, soit combiner le MDC et le chiffrement, soit alors se servir du MDC dans un canal authentifié. Les fonctions de hachage soupçonnées d'être à sens unique sont d'une grande importance dans la cryptographie car elles sont utilisées dans de nombreux algorithmes (MD2, MD4, MD5, SHA, etc.) et systèmes de cryptographie.

2.2.3 Génération de nombres aléatoires

La génération de nombres aléatoires intervient dans la production des clés de sessions. Une clé de session peut être perçue comme un identifiant de la connexion d'un ordinateur "client" à un ordinateur "serveur". La génération de nombres aléatoires peut aussi être utile à la création des vecteurs d'initilisation[5] d'un système crytographique, des secrets nécessaires à la production des signatures numériques, etc. Un générateur aléatoire est un dispositif capable de produire des nombres de façon aléatoire, imprévisible et non reproductible [99]. Un tel générateur est normalement doté d'un dispositif externe mesurant des phénomènes physiques connus par leur non-déterminisme ; c'est la cas par exemple d'une source radioactive ou quantique. La compagnie *id Quantique* s'est par exemple servi de la mécanique quantique [126] pour mettre sur pied un

[5]Il s'agit d'un ensemble de données nécessaire au démarrage d'un processus ; si un générateur aléatoire n'est pas disponible, la valeur d'une horloge ou un simple compteur peut généralement faire l'affaire.

générateur aléatoire, appelé *Quantis* [203].

Les générateurs pseudo-aléatoires sont des procédés déterministes développés à partir d'une séquence aléatoire initiale, qui peut être obtenue par diverses méthodes tels que la fréquence de frappe de l'utilisateur, le nombre d'accès disque, le nombre de paquets reçus par une interface réseau, etc. De nos jours, il existe de nombreux algorithmes implémentant les générateurs pseudo-aléatoires ; à ce propos, le lecteur peut par exemple consulter [99, 169].

Les primitives ci-dessus sont généralement utilisées dans des cryptosystèmes tels que RSA (des noms de ses auteurs **R**ivest, **S**hamir et **A**dleman) [146], AES (*Advanced Encryption Standard*) [71], PGP (*Pretty Good Privacy*) [185], etc. Elles permettent aussi de développer des protocoles pour sécuriser les transactions électroniques, sujettes à divers risques et attaques.

2.3 Attaques, Risques et Protection des données

Dans cette section, nous analysons les questions liées à la sécurité au niveau du client, du vendeur et des deux parties prises simultanément, dans un contexte de commerce électronique.

Le client doit être rassuré, par exemple, que le site Web commercial[6] appartient bien à une compagnie légitime ; que la page Web qu'il a devant lui n'a pas un contenu dangereux, plein de pièges ; que le site Web commercial ne va pas distribuer les informations qu'il s'engage à entrer à une tierce partie sans permission et, d'une manière générale, que ses échanges avec le site Web commercial ne vont pas nuire à sa vie privée.

Le vendeur quant à lui doit s'assurer que le client n'a pas pour objectif d'accéder à son serveur[7] pour modifier le contenu de ses pages et de son site Web, de rendre simplement indisponible son serveur, de disposer du contenu de son catalogue pour des besoins de compétitivité, etc. De manière générale, les sites Web commerciaux sont exposés à divers types d'attaques dont les plus fréquentes sont listées ci-dessous. Nous

[6] Boutique virtuelle (section 1.3.1) du vendeur.

[7] Machine qui stocke les données du site Web commercial. On parle aussi souvent de serveur Web.

resterons très sommaires, mais le lecteur peut avoir plus de détails en consultant [134, 173, 176, etc.]. On distingue généralement :

- Les attaques issues des failles liées aux systèmes d'exploitation, aux serveurs Web et de bases de données.
- Les attaques faites à travers le frontal (l'interface) du site de vente : le frontal peut en fait révéler des bogues du système ou du logiciel ou même de mauvaises configurations.
- Les attaques de denis de service distribuées ou DDoS (*Distributed Denial-of-Service*) qui visent essentiellement la saturation de toute la bande passante du site Web du vendeur. Cette catégorie inclut aussi les attaques de saturation de la mémoire : le malfaiteur envoie un gros volume de données au serveur pour occuper toute sa mémoire et provoquer par la même occasion l'exécution d'une requête dévastatrice contenue dans ces données.
- Les attaques par des données non valides entrées généralement sur la ligne contenant l'URL (*Uniform Resource Locator*). Ces attaques exploitent d'éventuelles failles issues du développement des scripts CGI (*Common Gateway Interface*) ou ASP (*Active Server Pages*), en vue de l'obtention du code source du script considéré et/ou du fichier des mots de passe.
- Les attaques dites de l'oreille indiscrète, à travers un canal de communication entre deux ordinateurs connectés ou non à Internet.
- Les attaques perpétrées à travers des virus généralement contenus dans du courrier électronique.
- L'attaque *Person-in-the-middle*, dans laquelle une entité C s'interpose entre deux entités A et B lors d'une transaction électronique. Cette attaque est expliquée et illustrée en section 2.4.8.2.
- L'*ingénierie sociale* est une technique qui consiste à obtenir un bien ou une information en exploitant la confiance, l'ignorance ou la crédulité de certains individus. Elle inclut le *hammeçonnage* ou *phishing*, qui consiste à obtenir des données confidentielles (mots de passe par exemple), en se faisant passer auprès des victimes pour quelqu'un digne de confiance.

– Les attaques par des *maliciels* : virus, vers, chevaux de Troie, etc.

Enfin, les deux parties voudraient s'assurer que le canal qui les connecte n'est pas rempli d'espions ou de malfaiteurs prêts à intercepter les informations échangées pour éventuellement en faire un autre usage. Elles voudraient également être sures que les informations échangées de part et d'autre ne sont pas modifiées en chemin.

Ces préoccupations nous conduisent à différents services de sécurité qu'il faut considérer dans les transactions en commerce électronique ; les plus courants [86, 118, 176] sont sont présentés dans la figure 2.1 et décrits ci-dessous :

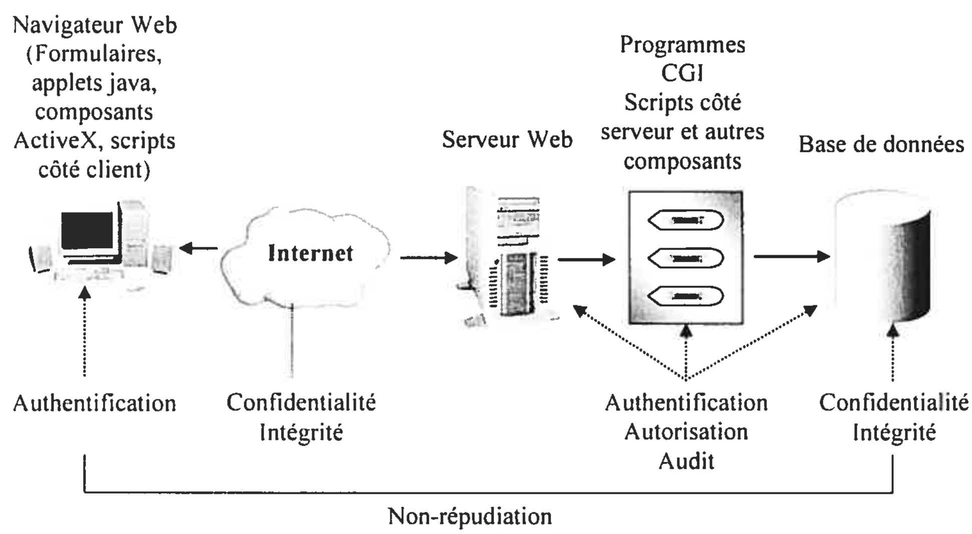


FIG. 2.1 – Services de sécurité dans le commerce électronique (*Inspirée de* [154]).

La confidentialité : c'est la protection de l'information contre une divulgation non autorisée. La confidentialité permet de protéger certains éléments sensibles tels que les contrats de ventes, les numéros de cartes de crédit, etc. Les attaques possibles sur ce service de sécurité sont l'analyse du trafic et les écoutes illicites. En cas de succès, ces attaques conduisent à la fuite d'informations. La confidentialité est mise en œuvre par les systèmes de chiffrement (section 2.4).

L'intégrité : c'est la protection de l'information contre une destruction ou une modification non autorisée ou accidentelle. En fait, il faut assurer l'intégrité des données pendant leur transit sur Internet et/ou après leur stockage. Les transactions financières constituent un exemple de données dont l'intégrité demande à être garantie. Les attaques qui visent l'intégrité consistent en la création, l'altération et la destruction illicite de l'information. L'un des moyens utilisés pour assurer l'intégrité des données est la combinaison "fonctions à sens unique + chiffrement", avant leur cheminement à travers des réseaux informatiques.

L'authentification : on parle généralement d'authentification d'entités et de celle de l'origine des données. La première assure la protection contre une fausse identité et donc, se confond à l'*identification*. Le vol de mots de passe est l'attaque la plus courante de cette première composante et, en cas de succès, cette attaque conduit à un accès non autorisé à l'information. Quant à la deuxième, elle donne une assurance sur la source des données et en assure l'intégrité. Ici, les risques sont la falsification et l'imitation des signatures numériques, avec pour conséquence possible la falsifiaction des données. Comme méthodes de protection, l'authentification d'entités peut, par exemple, nécessiter l'une des associations "secret[8] + protocole d'authentification", "adresse réseau + compte usager", "carte à puce + code PIN", etc. L'authentification des données peut, quant à elle, exiger la combinaison "fonctions à sens unique + chiffrement".

La non-répudiation : elle garantit le fait qu'une entité ne pourra pas nier d'être impliquée dans une transaction. L'un des moyens utilisés ici est la signature numérique. La prétention de vol de secrets est l'une des attaques majeures relatives à la non-répudiation. Cette attaque peut conduire à la contestation des transactions. Pour assurer la non-répudiation, on peut se servir de la combinaison "fonctions de hachage à collisions difficiles + chiffrement + signature numérique[9]".

[8]Exemples de secrets : clés secrètes, clés privées, etc.
[9]Voir, section 2.4.7.

La disponibilité : ce service donne l'assurance que les ressources (pages Web, données et autres services) sont accessibles aux utilisateurs légitimes. Les attaques possibles contre la disponibilité sont, par exemple, les tentatives d'accès pour briser le système et les attaques de denis de service. Ces attaques peuvent avoir pour résultats l'usage illicite du système, le denis de service, etc. Le contrôle d'accès et l'audit sont des exemples de méthodes de protection de la disponibilité.

L'anonymat : il table sur la préservation de l'identité d'une entité ou de la source d'une information ou d'une transaction. Dans la pratique, il est souvent mis en œuvre à travers l'utilisation des pseudonymes. L'analyse de trafic est un exemple d'attaque sur l'anonymat. Cette attaque peut avoir pour conséquence l'identification de l'entité (précédemment anonyme). Pour assurer l'anonymat, on peut avoir recours aux *mixeurs* et à la *monnaie électronique sans trace* [54].

L'autorisation : elle garantit que l'individu ou le programme a le droit d'accéder à une donnée particulière, à un programme ou à des ressources systèmes tels que les fichiers, les bases de registre, les répertoires, etc. Cette étape vient normalement après la phase d'authentification de l'entité, aussi appelée phase d'identification. L'autorisation tient souvent compte de l'individu (ou du programme) et des informations de contrôle d'accès associées à la ressource concernée (ex. : lecture seule, lecture/écriture, etc.).

L'audit : l'accès à un site Web conduit à l'écriture de certaines informations dans des fichiers log, au fur et à mesure que l'usager interagit avec le système. L'audit permet de revenir aisément sur les actions entreprises par divers usagers et d'identifier ceux qui sont à la base de mauvaises actions. Il est important de protéger les fichiers log contre toute suppression ou modification illicite. De tels fichiers doivent toujours être disponibles et exploitables.

En plus des services de sécurité présentés ci-dessus, les sites Web commerciaux sont généralement sécurisés au moyen des coupes feux (*Firewalls*), des réseaux virtuels privés ou VPN (*Virtual Private Networks*), des systèmes de détection d'intrusion ou IDS (*Intrusion Detection System*), etc.

Les coupes feux utilisés sont la plupart du temps de deux types : les filtres de paquets placés dans les routeurs et les serveurs *proxys* gérés par la couche application du protocole réseau TCP/IP [134]. Les VPN activent la sécurité des transmissions à travers l'Internet. Ils sont généralement utilisés comme support entre sites dans les transactions électroniques réunissant des partenaires du type de commerce électronique B2B (section 1.4). Ils sont aussi utilisés dans les communications entre un accès distant et un réseau local ou LAN (*Local Area Network*) central. En réalité, les transmissions Internet sont sécurisées en combinant le chiffrement, l'authentification et le contrôle d'accès entre les nœuds du réseau. Pour finir, les IDS [79] sont utilisés pour l'analyse des activités à travers le réseau local ou sur une machine hôte (distante) ; le but étant de découvrir toute activité suspecte et d'appliquer des actions automatisées appropriées s'il s'agit d'une attaque.

2.4 Cryptographie classique

Un cryptosystème classique consiste en trois composantes principales :

- un *algorithme de génération de clé* G : il permet de générer une clé de chiffrement, k, et une clé de déchiffrement, k'.
- un *algorithme de chiffrement* E : il utilise une clé k pour chiffrer un message m qu'Alice envoie à Bob et produit un cryptogramme, $c = E_k(m)$, qui circule dans un réseau susceptible d'être espionné.
- un *algorithme de déchiffrement* D : il permet à Bob de retrouver le message envoyé par Alice en se servant du cryptogramme c reçu et de sa clé secrète k'. Le mécanisme de déchiffrement se résume en : $m = D_{k'}(c) = D_{k'}(E_k(m))$.

On dira d'un cryptosystème qu'il est *symétrique* ou *asymétrique* selon que les clés k et k' utilisées respectivement dans le chiffrement et le déchiffrement sont égales ou distinctes. La figure 2.2 présente la vue globale d'un cryptosystème.

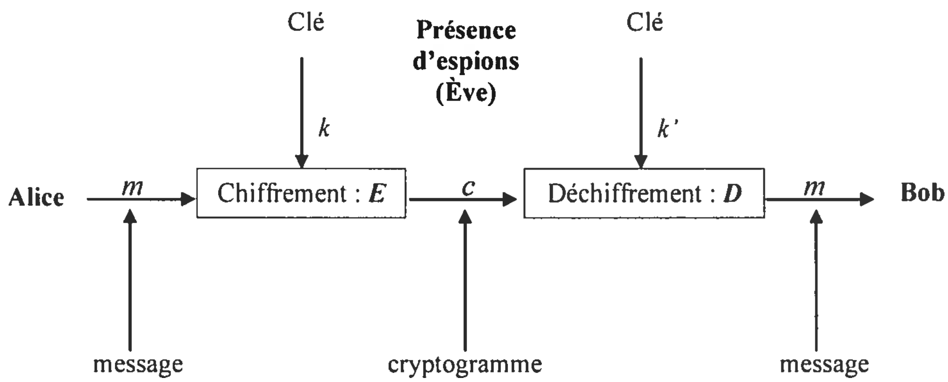


FIG. 2.2 – Cryptosystème.

La cryptologie est une discipline très complexe. Pendant que des cryptographes s'acharnent à concevoir de solides algorithmes de chiffrement ou de production de signatures numériques (Section 2.4.7), les cryptanalystes rassemblent des moyens humains et matériels (ex. : utilisation de plusieurs ordinateurs connectés à Internet) pour soumettre ces algorithmes à une rude épreuve de cryptanalyse. Cette section présente les deux types de cryptosystèmes classiques (symétriques et asymétriques), ainsi que les attaques possibles sur ces cryptosystèmes. Il y est aussi question de préciser comment choisir un cryptosystème pour des besoins de confidentialité. Nous mettrons fin à la section avec la classification des cryptosystèmes et les notions de signatures numériques, tiers de confiance et calcul multi-partie.

2.4.1 Systèmes symétriques

Les systèmes symétriques portent aussi le nom de cryptographie conventionnelle ou à clé secrète. L'un des plus anciens systèmes symétriques a été développé par Jules César (100-44 av. JC). L'idée de base est qu'à partir d'une seule clé sécrète, on puisse réaliser une transformation capable à la fois de rendre inintelligible et restituer une pièce d'information. Une telle transformation est illustrée dans la figure 2.3.

De nos jours, il existe de nombreux systèmes symétriques. Ils se subdivisent en deux groupes selon qu'ils utilisent le chiffrement par blocs (*block ciphers*) ou le chiffrement par flux (*stream ciphers*).

Dans le chiffrement par blocs, le texte est découpé en blocs, puis chiffré bloc après bloc. Les principaux algorithmes [168, 169] de ce groupe sont : DES (*Data Encryption Standard*), triple-DES (trois fois DES), AES (*Advanced Encryption Standard*), IDEA (*International Data Encryption Algorithm*), etc.

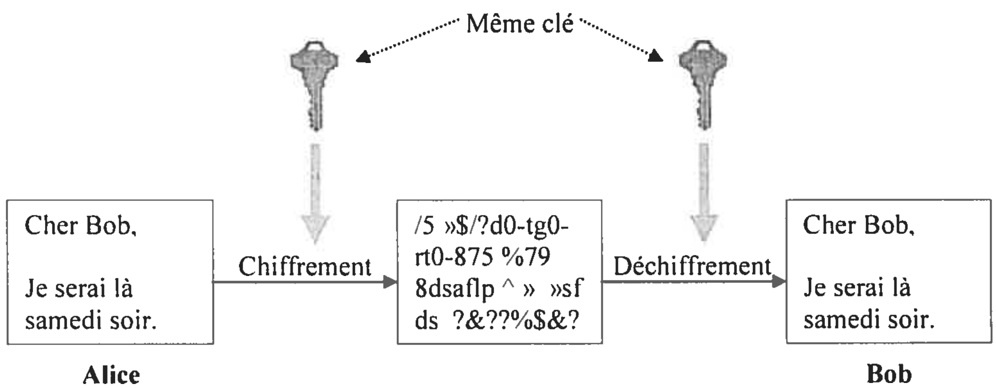


FIG. 2.3 – Cryptographie symétrique.

Le chiffrement par flux se ramène au traitement des données par unité de bit, à partir d'un vecteur d'initialisation. Les algorithmes les plus récents de ce groupe sont SNOW 2.0 [76], MUGI [181] et tous les algorithmes de chiffrement par blocs de 1 bit.

La présentation détaillée des méthodes de chiffrement par blocs ou par flux ne sera pas abordée ici, mais de nombreux auteurs traitent de l'un ou l'autre système [39, 99, 118, 168, 169, etc.]. En particulier, le lecteur peut consulter [92], qui est une encyclopédie récente sur le cryptographie et la sécurité.

La faiblesse majeure d'un système symétrique est le fait que sa sécurité réside entièrement dans le secret de la clé qui nécessite d'être transmise à travers un canal sécurisé. Sa force est qu'il est simple et facile à implémenter, en plus de sa rapidité en temps de calcul ; il est adapté au chiffrement de grands volumes de données. D'une manière générale, la cryptographie symétrique se prête mieux à la protection des documents personnels. Il faut souvent lui associer la cryptographie à clé publique (section 2.4.2) pour permettre l'échange de clés secrètes et, ainsi permettre la protection des données entre des participants éloignés.

2.4.2 Systèmes asymétriques

Depuis 1975 [10], la cryptographie classique a connu un véritable bond en avant avec la découverte des systèmes asymétriques [74,119] ou à clé publique. L'idée est d'utiliser deux clés différentes, l'une publique et l'autre privée, pour les opérations respectives de chiffrement et de déchiffrement. Pour recevoir des messages dans un système à clé publique, on doit donc disposer d'au moins une paire de clés : une clé publique qu'on met à la disposition de ses potentiels expéditeurs de messages et une clé privée qu'il faut garder *secrète*. L'envoi d'un message se fait au moyen de la clé publique du destinataire et ce dernier utilise sa clé privée pour le déchiffrement. Le chiffrement d'un message est relativement simple. En revanche, son déchiffrement est présumé difficile en l'absence de la clé privée. Les systèmes asymétriques sont illustrés dans la figure 2.4.

En outre, un cryptosystème à clé publique peut être *probabiliste*[11], auquel cas un ensemble, R, à partir duquel des *valeurs aléatoires* sont prises est requis. Ainsi, en plus de texte en clair et de la clé publique, l'algorithme de chiffrement prend en entrée un élément aléatoirement tiré de R pour produire le texte chiffré. Pour sa part, l'algorithme de déchiffrement n'a besoin que du texte chiffré et de la clé privée pour retrouver le texte en clair. Un des avantages des cryptosystèmes probabilistes est qu'ils aident à prévenir une attaque dans laquelle l'adversaire pourrait vérifier si un texte en clair donné est le bon en le chiffrant avec la clé publique et en comparant le texte chiffré obtenu avec celui qu'il compte décrypter. Cette attaque porte le nom de *guessing attack*. Le cryptosystème ElGamal [77] est un exemple de cryptosystème probabiliste (section 2.4.2.2).

Les systèmes asymétriques présentent un certain nombre d'avantages dont les plus importants sont l'aisance dans la distribution des clés publiques et l'existence de services

[10] La cryptographie à clé publique était connue depuis plusieurs années déjà.
[11] Il est à noter qu'un cryptosystème à clé secrète peut aussi être probabiliste.

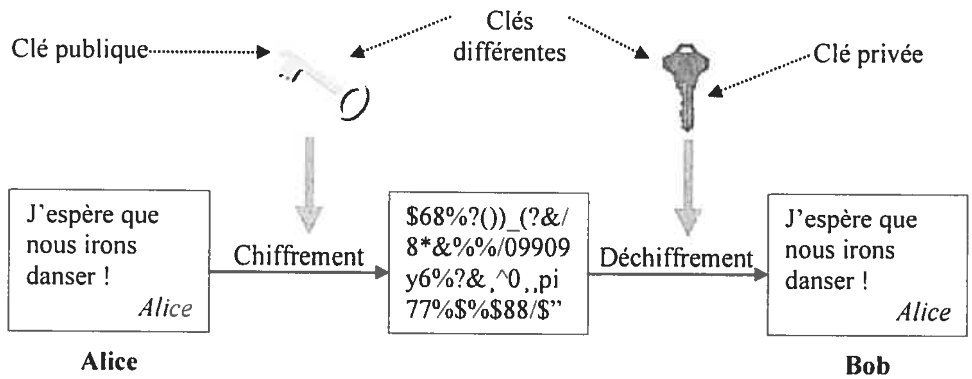


FIG. 2.4 – Cryptographie asymétrique.

telles que la confidentialité, l'authentification, l'intégrité, la signature numérique, la non-répudiation, etc.

En termes d'inconvénients, il y a une grande consommation de ressources en temps de calcul ; ils ne sont donc pas adaptés au chiffrement de flots de données importants. De plus, il faut gérer la publication et le cycle de vie des certificats de clé publique (section 2.4.8.3), ce qui n'est pas toujours chose facile.

Il existe de nos jours plusieurs systèmes de chiffrement asymétriques [169] (Rabin, ElGamal, etc.), mais le plus utilisé est RSA. La sécurité du système RSA [146] repose sur la difficulté présumée de factoriser un grand nombre, c'est-à-dire le décomposer en produit de nombres premiers. L'opération relativement simple avec de petits nombres est particulièrement difficile avec des nombres de plus de 200 positions décimales[12]. La clé privée dans RSA est fabriquée à partir de deux grands nombres premiers p et q alors que la clé publique ne révèle que leur produit $n = pq$. C'est le système le plus utilisé dans les transactions en commerce électronique.

Aux chapitres 4 et 5, certains protocoles que nous avons écrits utilisent le problème de Diffie-Hellman et le système de cryptographie à clé publique ElGamal. Aussi avons-

[12] Il est relativement facile de nos jours de factoriser des nombres de 200 positions décimales (660 bits), de sorte que la taille recommandée pour les clés RSA est désormais de 1024 bits.

nous jugé d'introduire ces deux notions ici.

2.4.2.1 Le problème de Décision Diffie-Hellman

Soient $\mathcal{G}$ un groupe et un élément quelconque g de $\mathcal{G}$. Étant donné un entier x, il est bien connu que le calcul de g^x dans $\mathcal{G}$ peut se faire de manière efficace avec un coût d'au plus $2 \lg x$ opérations de groupe en utilisant l'algorithme "*Square-and-multiply*" [169]. Toutefois, l'opération inverse, connue sous le nom de *Problème du Logarithme Discret* ou DLP (*Discrete Logarithm Problem*), est considéré comme étant difficile pour certains groupes bien choisis et pour des éléments g également bien choisis. De manière spécifique, le DLP consiste à trouver x tel que $a = g^x$, étant donné a et g dans $\mathcal{G}$.

Toutefois, en ce qui nous concerne, il ne sera pas suffisant de supposer que le DLP est difficile car nous basons nos protocoles sur le cryptosystème ElGamal (section suivante). Nous aurons plutôt besoin d'une hypothèse plus forte. Étant donné g^x, g^y et z dans $\mathcal{G}$, le problème de Décision Diffie-Hellman ou DDHP (*Decision Diffie-Hellman Problem*) consiste à décider si oui ou non $z = g^{xy}$. Le "Decision Diffie-Hellman *Assumption*"[13] ("DDH *Assumption*") stipule qu'il est difficile, dans certains groupes bien choisis, non seulement de résoudre le DDHP, mais aussi de distinguer entre le cas où $z = g^{xy}$ et celui où z est aléatoirement choisi dans $\mathcal{G}$, avec probabilité de succès significativement supérieure à $^1/_2$.

Considérons par exemple un grand nombre premier, p. Soient $\mathcal{G} = \mathbb{Z}_p^*$ le groupe multiplicatif des entiers modulo p (zéro étant exclu) et g un générateur de $\mathcal{G}$ (c'est-à-dire que chaque élement de $\mathcal{G}$ peut être obtenu par g^i pour i convenablement choisi, toutes les opérations sont effectuées dans $\mathcal{G}$ dans ce paragraphe). Dans ces conditions, pour p bien choisi, on pense que le DLP est difficile, alors qu'il est connu que le "DDH Assumption" n'est pas satisfaite. En effet, nous disons que $a \in \mathcal{G}$ est un *résidu quadratique* dans $\mathbb{Z}_p^*$ s'il existe $b \in \mathcal{G}$ tel que $a = b^2$. Cette propriété peut être vérifiée de manière efficace car $a \in \mathcal{G}$ est un résidu quadratique si et seulement si $a^{(p-1)/2} = 1$. Toutefois, il est facile de voir que g^{xy} est un résidu quadratique si et seulement si au moins g^x ou g^y est un résidu

[13]Cette expression ne se traduit pas bien en français ; nous la maintenons donc en anglais.

quadratique. À présent, étant donné g^x et g^y (quelconques), et un élément aléatoire z de $\mathcal{G}$, la probabilité que nous puissions affirmer que $z \neq g^{xy}$ est $^1/_2$, simplement parce que z n'est pas de la même résiduosité quadratique comparé à celle de g^x et g^y. Ce qui est suffisant pour invalider le "DDH Assumption" dans ce groupe.

Heureusement, le "DDH Assumption" est admis comme valide dans des groupes aussi simples que $\mathbb{Z}_p^*$. Considérons un grand nombre premier q tel que $p = 2q + 1$ soit aussi premier. Soit $\mathcal{G}$ le sous-groupe des résidus quadratiques dans $\mathbb{Z}_p^*$ ($\mathcal{G}$ contient en tout q éléments). Comme tous les éléments de $\mathcal{G}$ sont des résidus quadratiques, l'attaque sur le DDHP décrite ci-dessus ne tient plus la route. Jusqu'ici, il n'y a aucune autre attaque permettant d'invalider le DDHP dans ce groupe en supposant que le DLP est difficile.

2.4.2.2 Le cryptosystème ElGamal

Considérons un groupe cyclique multiplicatif $\mathcal{G}$ et un générateur g pour lequel nous supposons que le "DDH Assumption" est valide. Nous avons choisi l'ensemble $\mathcal{R}$ des entiers entre 1 et l'ordre de $\mathcal{G}$ (le nombre d'éléments dans $\mathcal{G}$) à la fois comme ensemble de valeurs aléatoirement choisies, ainsi que celui des clés secrètes.

Le cryptosystème ElGamal [77] est une système à clé publique probabiliste dans lequel l'espace des textes clairs $\mathcal{M}$ est $\mathcal{G}$ et l'espace des textes chiffrés $\mathcal{C}$ est $\mathcal{G} \times \mathcal{G}$.

- Algorithme de génération de clés : choisir $d \in \mathcal{R}$ de manière aléatoire. Prendre d comme clé privée et calculer la clé publique : $k = g^d$.
- Algorithme de chiffrement : Pour chaque clé publique k, la fonction de chiffrement $E_k : \mathcal{M} \times \mathcal{R} \to \mathcal{C}$ est définie par $E_k(m, r) = (g^r, m \cdot k^r)$.
- Algorithme de déchiffrement : Pour chaque clé publique k, nous rappelons que la clé privée correspondante est l'entier d tel que $k = g^d$. La fonction de déchiffrement $D_d : \mathcal{C} \to \mathcal{M}$ est définie par $D_d(e_1, e_2) = e_2 \cdot (e_1^d)^{-1}$.

Il a été démontré que briser le cryptosystème ElGamal (dans le sens strict de la sécurité sémantique[14]) est équivalent à briser le "DDH Assumption" [175]. Toute-

[14] Ici, l'adversaire est "passif" (il n'observe que des textes chiffrés) et a une puissance de calcul limitée. La sécurité sémantique veut qu'un tel adversaire ne soit pas capable d'obtenir de l'information significative sur le texte clair à partir du texte chiffré et de la clé publique.

fois, le détail d'implémentation devient important si nous voulons utiliser le groupe $\mathcal{G}$ des résidus quadratiques dans $\mathbb{Z}_p^*$, où $p = 2q + 1$ est tel qu'expliquer dans la section précédente. En effet, il ne serait pas pratique d'utiliser directement $\mathcal{G}$ comme espace de textes clairs $\mathcal{M}$ car, la moitié des éléments de $\mathbb{Z}_p^*$ seraient des textes clairs inexacts. Il serait plutôt intéressant de prendre $\mathcal{M}$ comme ensemble d'entiers entre 1 et q (inclusivement) et d'encoder $m \in \mathcal{M}$ en $u = m^2 \textbf{ mod } p$, considéré comme élément de $\mathcal{G}$, avant de le chiffrer avec E_k comme expliquer ci-dessus. Par ailleurs, après que u ait été retrouvé par l'application de D_d, il reste à extraire sa racine carrée modulo p. Heureusement, p est congruent à 3 modulo 4 car $p = 2q + 1$ et q sont impairs ; par conséquent, $(p + 1)/4$ est un entier. Dans ce cas, il est facile de vérifier que les deux racines carrées de u sont $\pm u^{(p+1)/4}$. Une fois les deux racines carrées réduites modulo p, exactement une seule parmi les deux sera comprise entre between 1 et q, et c'est l'élément correctement déchiffré $m \in \mathcal{M}$.

Dans notre modèle BCBB (chapitres 4 et 5), nous nous servons du cryptosystème ElGamal pour chiffrer les échanges entre Alice et Sir Bob.

2.4.3 Attaques sur les systèmes cryptographiques

Différents types d'attaques peuvent être faits sur des systèmes cryptographiques :

- **Attaque à texte chiffré seulement** (*ciphertext-only*) : l'adversaire essaie de trouver la clé ou les textes clairs à partir de l'observation de textes chiffrés dont les messages correspondants sont inconnus. L'adversaire suppose qu'une même clé inconnue à été utilisée. Un système cryptographique vulnérable à cette attaque n'offre aucune sécurité.
- **Attaque à texte clair connu** (*known-plaintext*) : l'adversaire a des couples (texte clair, texte chiffré) à sa disposition et possiblement d'autres textes clairs. Son objectif est alors de trouver la clé secrète.
- **Attaque à texte clair choisi** (*chosen-plaintext*) : l'adversaire peut choisir un ou plusieurs textes clairs et obtenir les textes chiffrés correspondants, le but étant de retrouver la clé secrète.
- **Attaque dynamique à texte clair choisi** (*adaptive chosen-plaintext*) : il s'agit

d'une attaque à texte clair choisi dans laquelle le choix du texte clair à une étape donnée peut dépendre des textes chiffrés reçus lors des étapes précédentes.

- **Attaque à texte chiffré choisi** (*chosen-ciphertext*) : l'adversaire choisit un ou plusieurs textes chiffrés et obtient les textes clairs correspondants. L'objectif de cette attaque est de trouver la clé.
- **Attaque dynamique à texte chiffré choisi** (*adaptive chosen-ciphertext*) : il s'agit d'une attaque à texte chiffré choisi dans laquelle le choix du texte chiffré à une étape donnée peut dépendre des textes clairs reçus lors des étapes précédentes.
- ***Boomerang attacks*** : c'est une nouvelle méthode d'attaque qui regroupe les attaques dynamiques à textes clairs choisis et à textes chiffrés choisis.

2.4.4 Confidentialité : quel système choisir ?

La confidentialité est assurée par le chiffrement. Il existe des centaines d'algorithmes symétriques et asymétriques capables de fournir un niveau de confidentialité suffisant. Les solutions symétriques offrent des avantages en terme de rapidité, de facilité d'implémentation sur le matériel, de longueur de clé réduite (en moyenne 128 bits, soit 16 caractères ASCII). Les solutions asymétriques quant à elles simplifient l'échange et la gestion des clés. En effet, l'échange se fait à travers un canal authentifié mais non confidentiel et, en terme de gestion de clés, une seule clé publique[15] suffit à un utilisateur pour recevoir des messages confidentiels de n utilisateurs, ce qui nécessiterait n clés différentes dans le cas symétrique.

Il existe au moins deux problèmes communs aux deux systèmes : la gestion de clés par l'utilisateur et les restrictions légales d'usage et d'exportation ; par exemple, les États-Unis interdisent l'exportation de la cryptographie dite *forte*. Dans ces conditions, quel système faut-il utiliser, symétrique ou asymétrique ? La réponse à cette question dépend de ce qui nécessite d'être protégé. Le tableau 2.1 illustre quelques cas typiques d'utilisation de la cryptographique symétrique et asymétrique. En particulier, il est parfois utile de penser à des systèmes cryptographiques mixtes ; l'idée étant d'utiliser la

[15] L'utilisateur garde secrète la clé privée associée à cette clé publique.

TAB. 2.1 – Cryptographie symétrique versus cryptographie asymétrique

Activité	**Recommandation**	**Remarques**
Protection de documents personnels	Cryptographie symétrique	Rapidité et clés faciles à mémoriser, car plus courtes.
Protection de documents dans un groupe d'utilisateurs proches	Cryptographie symétrique	Rapidité et facilité d'échange des clés confidentielles.
Etablissement de canaux confidentiels entre utilisateurs distants (inconnus)	Cryptographie asymétrique	Nul besoin d'avoir un canal confidentiel : un canal authentifié suffit.
Transactions entre deux utilisateurs distants	Cryptographie asymétrique pour la protection de la clé symétrique et cryptographie symétrique pour la protection des données	Rapidité et changement possible de la clé symétrique pour chaque correspondant.

cryptographie publique uniquement pour échanger des clés symétriques. Si Alice et Bob veulent faire des échanges, Alice génère une clé aléatoire k et la transmet à Bob en la chiffrant avec la clé publique de Bob, ce qui assure la confidentialité de k. Bob déchiffre au moyen de sa clé privée et obtient la clé k. À partir de cet instant, Alice et Bob sont authentifiés (l'un vis-à-vis de l'autre) et communiquent en utilisant la clé commune k pour assurer la confidentialité de leurs échanges.

2.4.5 Authentification de l'origine des données

Supposons que l'on ait deux individus, Alice et Bob, qui communiquent. Comment faire pour rassurer Bob que les données, en l'occurrence les messages, qu'il reçoit proviennent effectivement d'Alice ? Ce problème porte le nom d'authentification de l'origine des données et nous présentons ci-dessous trois méthodes qui peuvent le résoudre. (Les notions de *MAC* et *MDC* utilisées ici ont été introduites en section 2.2.2.)

La première possibilité consiste à se servir du *MAC*, avec une clé secrète k connue d'Alice et Bob. Alice envoie à Bob à la fois le message m et le digest $MAC_k(m)$, c'est-à-dire le MAC appliqué au message m en se servant de la clé secrète k. Dès la réception de m et du digest, Bob recalcule de son coté $MAC_k(m)$ à partir de m et de la clé secrète k. S'il obtient la même valeur ($MAC_k(m)$ envoyé par Alice), il conclut que le message provient d'Alice.

La deuxième possibilité combine le MDC et le chiffrement symétrique utilisant une clé secrète, k, connue d'Alice et Bob. Alice envoie à Bob le message m et le cryptogramme, $E_k(MDC(m))$, issu du chiffrement du digest. Puis, Bob calcule $MDC(m)$ à partir de m et $E_k(MDC(m))$ à partir de sa connaissance de k. S'il obtient la même valeur ($E_k(MDC(m))$ envoyé par Alice), il conclut que le message provient d'Alice. On peut faire mieux en rendant le message m confidentiel. Plus concrètement, Alice envoie à Bob le cryptogramme issu du chiffrement du digest (application du MDC au message) et du message : $E_k(m, MDC(m))$. Bob procède donc d'abord au déchiffrement du cryptogramme reçu, en utilisant la clé secrète k. Il obtient le message m et le digest $MDC(m)$. Il vérifie alors si le digest reçu correspond bien au message m en appliquant MDC à m et en comparant avec le digest résultant du déchiffrement.

La dernière possibilité est la combinaison du MDC et d'une signature numérique (section 2.4.7). En effet, la notion de signature numérique fait appel aux cryptosystèmes à clés publiques que nous avons présentés en section 2.4.2. Nous supposons donc qu'Alice dispose d'une clé privée S_A et que la clé publique P_A correspondante a été mise à la disposition de Bob. Ainsi, pour authentifier un message m, Alice envoie m et $SIG_{S_A}(MDC(m))$ à Bob, où SIG_{S_A} renvoie à l'application de la signature au moyen de la clé privée S_A. Bob calcule alors $MDC(m)$ et compare $SIG_{S_A}(MDC(m))$ avec une copie authentique de P_A. En cas d'égalité, Bob conclut qu'Alice est à l'origine du message. Cette solution offre en plus la non-répudiation de l'origine du message, c'est-à-dire, Alice ne pourra pas nier avoir envoyé le message m à Bob, sauf si elle prétend qu'on lui a volé sa clé privée ou que son ordinateur était infecté par un virus et qu'elle a plutôt signé un autre message apparu sur son écran.

Les protocoles ci-dessus sont simples mais n'offrent aucune indication sur l'unicité et l'actualité des messages reçus. Ils nécessitent des mécanismes tenant compte du temps ou du contexte de la transaction. Il faut donc leur associer des mécanismes d'authentification d'entités, encore appelée identification (section 2.3). Alice peut par exemple se servir des éléments suivants pour prouver son identité : les choses qu'elle connaît (mots de passe, codes PIN, clés privées ou secrètes, etc.), les choses qu'elle possède (passeport, carte à puces, générateurs de mots de passe, etc.) et les choses inhérentes à un être humain (propriétés biométriques tels que les empreintes digitales, la rétine, le code ADN, etc.).

2.4.6 Classification des systèmes cryptographiques

Un système cryptographique peut être inconditionnellement sécuritaire (du moins, on l'espère !), calculatoirement sécuritaire ou à sécurité « démontrable ».

2.4.6.1 Sécurité inconditionnelle

Un système cryptographique[16] est dit inconditionnellement sécuritaire (*unconditionally secure*) si sa sécurité n'est pas compromise par la puissance de calcul utilisée par la cryptanalyse. Cette catégorie s'appuie sur la théorie de l'information de Shannon [160]. Plus précisément, un système cryptographique est inconditionnellement sécuritaire si la probabilité d'obtenir un texte en clair, m, après l'observation du texte chiffré correspondant, c, est identique à la probabilité *a priori* d'obtenir le texte en clair, m. En d'autres termes, le fait de disposer de couples formés des textes clair et chiffré, (m, c), ne constitue aucune aide pour la cryptanalyse. Une condition nécessaire pour qu'un système soit inconditionnellement sécuritaire est que la clé de déchiffrement soit au moins de la même *entropie* que le message et, surtout, qu'elle ne soit pas réutilisée pour chiffrer des messages différents. Cette condition rend ces systèmes peu adaptés aux besoins cryptographiques contemporains et réduit leur domaine d'intérêt à un cadre théorique. L'exemple classique est le *masque jetable* (*One-time pad*) inventé[17] par Joseph Mauborgne and Gilbert Vernam en 1917. Le principe du masque jetable consiste à chiffrer chaque bit du texte clair avec un bit de clé, de sorte que la clé secrète est aussi longue que tout le texte

[16]On définit de façon similaire la sécurité inconditionnelle pour un système d'authentification.
[17]Rivest [144] attribue l'invention à Vernam alors que Singh [164] l'attribue à Mauborgne.

clair à chiffrer.

2.4.6.2 Sécurité calculatoire

Un système de cryptographie est dit calculatoirement sécuritaire (*computationally secure*) si l'effort de calcul nécessaire pour le casser en utilisant les meilleures techniques publiquement connus est au-delà des ressources de calcul dont peut disposer l'adversaire. La grande majorité des systèmes cryptographiques symétriques (AES, DES, IDEA, etc.) et asymétriques (RSA, ElGamal, etc.) offrent une sécurité calculatoire, mais on n'est jamais très sûr.

2.4.6.3 Sécurité « démontrable »

Un système de cryptographie est dit à sécurité démontrable (*demontrably secure*) si on peut prouver que le fait de réussir à le cryptanalyser est au moins aussi difficile que de résoudre un problème générique de base réputé difficile. Comme exemples de problèmes génériques, on peut citer, entre autres, la factorisation de grands nombres, le calcul des racines carrées modulo un nombre composé et le calcul de logarithmes discrets dans un groupe fini. Les systèmes qui satisfont à ce critère sont rares ; le plus connu reste celui de Rabin [136] dont on peut démontrer l'équivalence à la factorisation pour une attaque à texte chiffré seulement.

2.4.7 Signatures numériques

Les signatures numériques permettent d'arbitrer les disputes relatives à l'émission de messages numériques ; par exemple, ils jouent un rôle important dans la base de sécurité "non-répudiation" (sections 2.3 et 2.4.5). Une signature numérique est une chaîne de données permettant d'associer un message, sous forme numérique, à l'émetteur dudit message. Le schéma de signature numérique est l'ensemble formé d'un algorithme de génération de signature et d'un algorithme de vérification de signature.

On distingue deux classes de signatures numériques. D'une part, les signatures qui nécessitent la présence du message original pour vérifier la validité de la signature. Elles sont les plus couramment utilisées. La *signature ElGamal* [77] et l'agorithme DSA (*Digital Signature Algorithm*), bâti sur le standard DSS (*Digital Signature Standard*) [80], sont des exemples de protocoles de signature numérique. D'autre part, on a les signatures numériques avec reconstitution du message qui offrent, en plus, la possibilité de reconstruire le message à partir de la signature. C'est le cas des signatures numériques RSA [146].

Les protocoles de signature numérique sont pour la plupart basés sur la cryptographie asymétrique. Des engagements semblables à ceux obtenus par une signature à clé publique, comme la non-répudiation de l'origine du message, peuvent cependant être obtenus avec les systèmes symétriques et des tiers de confiance ou TTP (*Trusted Third Parties*) que nous présenterons en section 2.4.8.

La signature de tout un document est souvent très coûteuse en temps. Il faut donc uniquement faire la signature du digest résultant de l'application d'une fonction de hachage à sens unique sur tout le document. L'algorithme de signature utilise la clé privée du signataire. La figure 2.5 illustre le schéma de génération de signature.

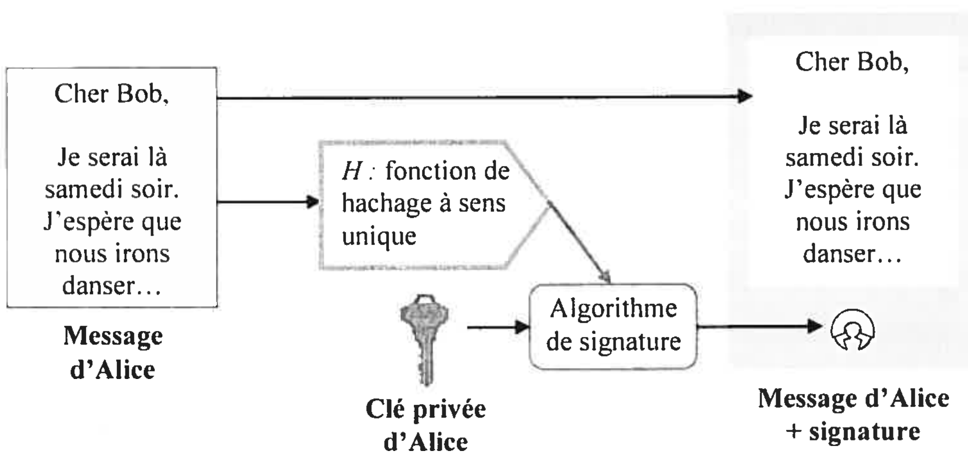


FIG. 2.5 – Exemple de génération de signature.

L'étape de vérification de signature suppose, d'une part, l'application de la fonction de hachage à sens unique (identique à celle utilisée pendant la génération de la signa-

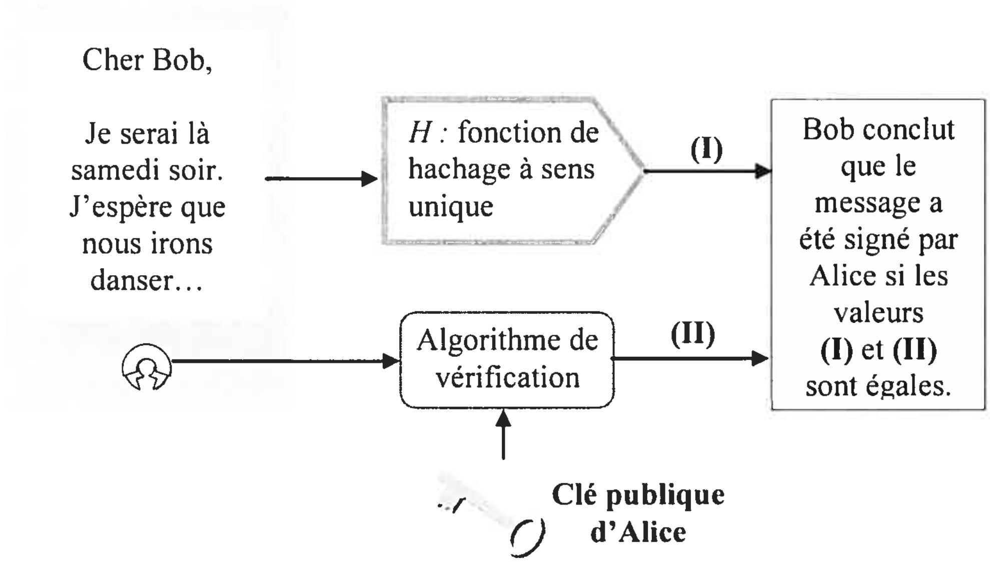


FIG. 2.6 – Exemple de vérification de signature.

ture) sur le message et, d'autre part, l'application de l'algorithme de vérification sur la signature qui accompagne le message. La signature est valide si dans les deux cas, on a la même valeur. L'algorithme de vérification utilise la clé publique du signataire. La figure 2.6 représente le schéma de vérification d'une signature.

Les caractéristiques d'une signature numérique sont généralement les suivantes :

- la signature change si le document change, mais les clés publique et privée sont toujours les mêmes ;
- en cas de modification du document ou de la signature, la signature ne sera pas valide, ce qui garantit l'intégrité ;
- il est difficile, même pour le détenteur de la clé privée, de générer un deuxième document avec la même signature ;
- seul le détenteur de la clé privée peut générer une signature qui se vérifie avec la clé publique correspondante.

Toutes ces caractéristiques garantissent donc les services d'authenticité et la non répudiation.

Signatures aveugles : En 1982, Chaum [55] a mis au point un schéma de signature numérique dénommé signatures aveugles (*blind signatures*). Son fonctionnement est tel que si Alice envoie une information à Bob pour en obtenir sa signature, Bob retourne à Alice l'information signée. À partir de cette signature, Alice peut calculer la signature de Bob sur un autre message qu'elle a choisi à l'avance. Ceci permet à Alice d'avoir une signature de Bob sur un message que B n'a jamais vu : d'où le nom de *signature aveugle*.

Attaques sur les signatures numériques : Un schéma de signature numérique peut être "brisé" de trois façons :

- Bris total : l'adversaire calcule la clé privée du signataire ou trouve un algorithme efficace, c'est-à-dire fonctionnant en temps polynomial, pour générer des signatures.
- Falsification sélective : l'adversaire est capable de générer une signature valide pour un message ou une classe de messages précis(e).
- Falsification existentielle : l'adversaire est capable de contrefaire une signature pour au moins un message sur lequel il n'a pas le contrôle. C'est cette falsification que Chaum a exploité dans le cas de RSA pour créer des signatures aveugles.

2.4.8 Tiers de confiance

Un tiers de confiance ou TTP (*Trusted Third Party*) est une entité qui joue en quelque sorte le rôle d'arbitre dans des échanges entre deux ou plusieurs parties. Il a trois modes de fonctionnement : *in-line*, *on-line* et *off-line*. Supposons qu'Alice et Bob font des échanges via un TTP. Dans le mode in-line, le TTP agit comme un intermédiaire pour relayer en temps réel les échanges entre Alice et Bob. C'est le cas des serveurs proxys et des passerelles sécurisées (*Secure Gateways*) [134]. En mode on-line, le TTP participe en temps réel aux échanges entre Alice et Bob qui, toutefois, communiquent directement (sans passer par le TTP) ; c'est le cas par exemple des centres de distribution de clés. Pour finir, en mode off-line, le TTP ne participe pas à l'échange en temps réel mais rend l'information disponible *a priori*. C'est le cas des autorités de certification. Contrairement aux deux premiers, en mode off-line, les échanges sont facilités et on n'a pas besoin de la disponibilité permanente du TTP.

Les TTP sont particulièrement sollicités dans les échanges faits sur Internet en général et, en particulier, dans la pratique du commerce électronique. Nous donnons ci-dessous quelques exemples de TTP.

2.4.8.1 Les centres de distribution de clés

Dans un environnement symétrique de n entités sans intermédiaire, $n(n-1)/2$ (soit environ n^2) clés différentes sont nécessaires pour que toutes les paires d'entités partagent une clé différente. De plus, un tel système n'est pas évolutif car l'ajout d'une entité se traduit par la génération de n nouvelles clés. Le but d'un centre de distribution de clés ou KDC (*Key Distribution Centre*) est de résoudre le problème de distribution de clés. En effet, si chaque entité partage une clé avec un KDC, seules n clés sont nécessaires pour le fonctionnement du système et une clé suffit pour chaque nouvelle entité. L'établissement de canaux sûrs est quant à lui assuré, entre autres, par la génération de clés de session. Il y a toutefois un certain nombre de problèmes. D'abord, si le KDC est compromis, à travers par exemple l'usurpation d'identité des entités faisant partie du réseau d'échange, alors tout le système devient vulnérable. De plus, le mode de fonctionnement habituel d'un KDC étant on-line, s'il devient indisponible suite par exemple à une attaque de denis de service [173] (section 2.3), alors tout le système est paralysé. Enfin, les opérations des KDC sont souvent coûteuses en temps de calcul (chiffrement/déchiffrement, génération aléatoire et pseudo-aléatoire, etc.).

2.4.8.2 Les autorités de certification

Une autorité de certification ou CA (*Certification Authority*) a pour rôle d'authentifier l'association entre une entité donnée et sa clé publique. La CA crée et signe des certificats contenant cette association (moyennant une preuve d'identité comme un passeport) et les rend accessibles aux entités qui les ont sollicités. Une fois signées, des copies de certificats peuvent être gardées dans des endroits non protégés, par exemple dans l'espace disque de l'utilisateur. Cependant, afin de vérifier la signature des certificats, il y a la nécessité de disposer d'une copie authentique de la clé publique de la CA. On n'a pas besoin d'implanter des protocoles complexes dans une CA. Le mode de fonctionnement habituel d'une CA est off-line, ce qui diminue les conséquences liées à des périodes d'indisponibilité. Les CA publient des listes signées des certificats non valides ou CRL (*Certificate Revocation Lists*). En réalité, il s'agit de listes contenant des certificats devenus non valables suite à une clé privée compromise ou à tout autre facteur justifiant la non-validité des informations contenues dans un certificat (changement de l'algorithme utilisé, etc.). Une CA se doit de publier des CRL avec une fréquence très élevée et en utilisant des canaux de distribution de large audience, afin de diminuer le risque de fraudes.

Avec une CA, on peut par exemple éviter une attaque de type *Person-in-the-middle* illustrée dans la figure 2.7, dans laquelle, une espionne, Ève, se glisse entre deux correspondants pour substituer les messages légitimement échangés. Pour illustrer ce type d'attaque, nous supposons qu'Alice est un utilisateur Web, potentiellement un client de commerce électronique, contrairement à Bob qui dispose d'un site Web, en l'occurence un site Web commercial. Le rôle de Ève est donc de faire croire à Alice et Bob qu'ils font des échanges, alors qu'en réalité, Ève a créé une "Alice fictive" pour Bob et un "Bob fictif" pour Alice. Ève envoie alors ce qu'elle veut de chaque côté et fait une copie des données pertinentes tels que les numéros de cartes de crédit, les clés d'activation de logiciels et de films, etc.

Le cas Alice/Bob sur le Web que nous venons de présenter trouve la solution dans la certification. En effet, si le site Web de Bob dispose d'un certificat d'authentification, alors Alice va apprendre le nom certifié de ce site Web.

D'une manière générale, voici quelques éléments qui entrent dans la composition d'un certificat d'authentification :

- Le numéro de série et la version.
- L'émetteur : l'identité de la CA signataire.
- L'algorithme de signature : l'algorithme permettant de calculer la signature sur le

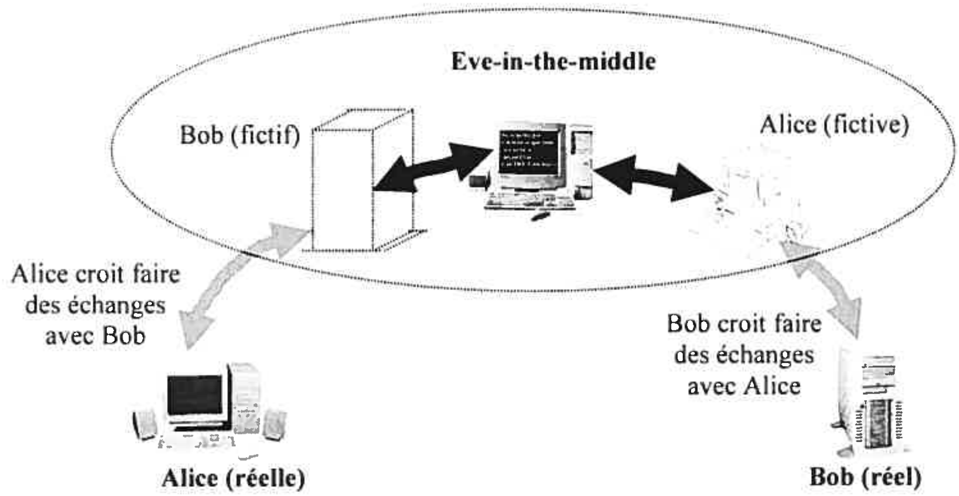


FIG. 2.7 – Attaque Person-in-the-middle.

certificat, par exemple, les combinaisons MD5 et ElGamal, SHA-1 et RSA, etc.

- Le destinataire : le nom de l'entité dont la clé publique est certifiée.
- La clé publique de l'entité destinataire.
- L'algorithme de chiffrement associé à la clé publique (ex. : RSA)
- La validité : la période de validité du certificat.
- La signature : ce champ contient la signature effectuée au moyen de l'algorithme de signature et de la clé privée de la CA. Elle porte sur l'ensemble des éléments précédents et garantit ainsi l'authenticité des informations qu'ils contiennent.

2.4.8.3 Infrastructure à clé publique

Une infrastructure à clé publique ou PKI (*Public Key Infrastructure*) [3, 100, 178] est une infrastructure intégrée permettant de fournir un ensemble de services de sécurité sur la base de la cryptographie à clé publique. Les fonctionnalités d'une PKI sont les suivantes :

- La CA pour la création et la maintenance des certificats.
- Le répertoire des certificats pour la mise des certificats à la disposition des utilisa-

teurs et des applications.

- La révocation des certificats compromis ou devenus obsolètes.
- La sauvegarde et le rétablissement centralisés des clés ; ceci permet de gérer la perte de clés suite à des événements divers (destruction du support matériel, oubli du mot de passe de déblocage, départ de l'employé, etc.).
- La mise à jour automatique des clés après la fin de leur validité.
- L'historique des clés et des certificats pour la récupération des clés devenues obsolètes mais ayant servi à chiffrer un document dans le passé.
- La certification croisée avec d'autres PKI (clients, fournisseurs, partenaires, etc.). Cette fonctionnalité peut permettre de valider les certificats émis par d'autres PKI.
- Le support pour la non-répudiation : c'est un service à valeur ajoutée permettant de fournir certaines évidences telle que l'authentification de l'origine des données lors du déroulement d'une transaction.
- Le logiciel client qui permet de réaliser toutes les opérations propres à la PKI côté client, en l'occurrence la gestion des certificats utilisateurs, la signature de documents, le déchiffrement d'information, la gestion de périphériques spécifiques (lecteurs de cartes à puces, dispositifs biométriques, etc.).

Les avantages d'une PKI sont très nombreux. Nous en donnons les trois principaux. Il y a *d'abord* l'avantage lié à la sécurité : la nature intégrée d'une PKI permet de créer un environnement de sécurité sans maillons faibles. *Ensuite*, une PKI est un tout-en-un puisqu'elle permet l'intégration et la gestion de tous les paramètres de sécurité propres à un grand nombre de services tels que l'authentification d'entités et la signature numérique de documents. Ces deux services donnent naissance à la non-répudiation, aux réseaux privés virtuels, à des communications sécurisées entre clients, partenaires, fournisseurs des types de commerce électronique B2C et B2B, etc. Il faut *enfin* noter l'interopérabilité intra et inter compagnie : les principaux produits d'une PKI répondent à des normes et standards [201] très répandus (ex. : X.509). Un grand nombre d'applications et de dispositifs matériels sont désormais conformes à ces standards.

Pour ce qui est des inconvénients, on peut essentiellement parler du coût de mise en place : l'insfrastruture coûte cher et les compétences sont rares.

Pour terminer, d'autres types de TTP existent, parmi lesquels se retrouvent les agents d'horodateur ou TAs (*Timestamp Agents*) [197] chargés de certifier l'existence d'un document ou le déroulement d'une transaction à un moment bien spécifié dans le temps ; les agents notaires ou TA (*Notary Agents*) [142] qui certifient non seulement l'existence d'un document à un temps donné (comme les TA), mais également sa validité, son origine ou son appartenance à une entité donnée. Ce service peut constituer un support légal pour la non-répudiation car il passe pour un acte notarié.

2.4.9 Calcul multi-partie

Un cadre général pour présenter des problèmes cryptographiques [84] consiste à spécifier un processus qui envoie n entrées $(x_1, \cdots, x_n)$ vers n sorties $(y_1, \cdots, y_n)$. Les entrées d'un tel processus peuvent être vues comme les entrées locales de n participants, et les n sorties sont leurs sorties locales correspondantes. Le processus décrit la fonction désirée. C'est-à-dire, si les n parties font mutuellement confiance à une $(n+1)$-ième partie externe, alors elles peuvent toutes envoyer leurs entrées locales à cette partie, qui pourrait calculer la sortie du processus et retourner à chacune d'entre-elles la sortie correspondante. Une question essentielle dans ce cas est comment mettre à la disposition des parties, qui mutuellement ne se font pas confiance, une tierce partie virtuelle ?

Le calcul multi-partie ou MPC (*Multi-Party Computation*) répond à cette question. En effet, le MPC concerne l'étude générale des calculs sécurisés entre n participants $P_1, P_2, \cdots, P_n$ reliés par des canaux sécuritaires ; chaque participant P_i dispose d'une entrée x_i et le but est de calculer une valeur $f(x_1, x_2, \cdots, x_n)$ pour une certaine fonction, f, connue de tous. Le MPC exige la prise en compte du nombre de participants corrompus, ainsi que des hypothèses sur les canaux reliant les participants.

Plus précisément, les participants corrompus ne doivent rien apprendre autre que de la sortie de la fonction f. Aussi, ils ne doivent pas affecter la sortie, sauf peut-être par le choix de leurs entrées. En outre, les modèles de communication diffèrent en fonction de trois critères : l'existence de canaux secrets de communication reliant chaque paire de participants, l'existence d'un canal de diffusion qui permet de transmettre le même message à plusieurs destinataires en même temps, et suivant des contraintes temporelles sur les canaux de communication synchrones (un message émis arrive immédiatement au(x) destinataire(s)) ou asynchrones (existence d'un délai de transit variable selon le moment d'émission ou la destination).

Le problème du calcul à plusieurs parties a été étudié pour la première fois par Yao [184]. La première solution générale pour ce problème a été proposée la même année et de manière indépendante par Chaum, Damgård et de Graaf [58] d'une part et, d'autre part, Goldreich, Micali et Wigderson [88]. Par la suite, de nombreux résultats ont suivi [28, 29, 57, etc.].

Si on a $n = 2$, on parle de *calcul securitaire à deux participants* ou STPC (*Secure Two Party Computation*). Dans ce cas, deux participants, Alice et Bob, désirent évaluer une fonction $f(x, y)$ avec des entrées x et y fournies respectivement par Alice et Bob, de sorte que, au terme de l'évaluation, Alice et Bob apprennent le résultat mais demeurent ignorants respectivement des valeurs y et x. Le premier protocole de STPC a aussi été introduit par Yao [184]. Dans un contexte de commerce électronique, le STPC est matérialisé dans les échanges entre le client et le vendeur au cas où ils voudraient respectivement protéger la vie privée et les données sensibles, comme c'est le cas dans cette thèse.

2.5 Cryptographie et commerce électronique

Deux principaux protocoles [18, 86, 154, 176] de sécurité ont été mis au point pour garantir la sécurité du commerce électronique dans la gestion des transactions et le chiffrement des communications. Le protocole SSL (*Secure Socket Layer*) permet le chiffrement des transactions électroniques.

Nous allons faire une étude des différents protocoles cryptographiques utilisés en commerce électronique [86, 154, 179].

2.5.1 Le protocole SSL

Le protocole SSL (*Secure Sockets Layer*) a été conçu et implémenté par Netscape et on le situe entre la couche transport (TCP) et les protocoles de la couche application (HTTP, SMTP, FTP, etc.) du modèle TCP/IP [134]. Il offre les services de confidentialité, d'intégrité, d'authentification du serveur et, accessoirement, du client. Le SSL est basé sur des algorithmes de cryptographie à clé publique, en l'occurrence, RSA [146] et Diffie-Hellman [74]. L'intervention des CA pour certifier l'association entre les entités et les clés publiques est vivement recommandée mais pas indispensable [99]. Le SSL 3.0 est la version courante et demeure une technologie propriétaire (Netscape). Une version équivalente a été adopté par l'IETF (*Internet Engineering Task Force*) et est connue sur le nom de TLS 1.0. Une version plus récente, TLS 1.1, est couramment utilisée de nos jours. Le SSL est une "mini-pile" de protocoles avec des fonctionnalités pour les couches session, présentation et application du modèle OSI[18]. Il est constitué de trois blocs fondamentaux : le SSL *record protocol*, qui permet l'encapsulation des protocoles de plus haut niveau au-dessus de TCP (fragmentation, compression et chiffrement) ; le SSL *handshake protocol* qui est chargé de l'authentification des intervenants et de la négociation des paramètres de chiffrement ; et le SSL *state machine*. Pour ce dernier bloc, SSL est un protocole à états. Il nécessite donc un ensemble de variables qui déterminent l'état d'une session et d'une connexion [99]. Contrairement à SSL 3.0, TLS opère au niveau de la couche transport du modèle OSI et peut être utilisé dans des trafics autres que ceux du HTTP. Par exemple, si on ne sert pas d'un protocole HTTP, l'accès sécuritaire aux courriels peut se faire en utilisant TLS.

[18]Le modèle OSI (*Open Systems Interconnection*) est le modèle de référence d'interconnexion de systèmes ouverts, créé par l'ISO (International Organization for Standardization, `www.iso.org`) afin d'offrir une base commune à la description de tout réseau informatique. [*Source :* `www.wikipedia.org`]

2.5.2 Le protocole PGP

PGP (*Pretty Good Privacy*) est une technologie développée en 1991 par Phil Zimmerman [185] qui l'a distribuée gratuitement sur Internet[19]. Cette technologie fournit les services d'encapsulation des données, de gestion et d'authentification des clés. Il demeure gratuit et de libre accès jusqu'ici. PGP fournit des services de sécurité au-dessus du protocole de gestion et transfert de courrier électronique SMTP (*Simple Mail Transfer Protocol*) et il n'affecte que le contenu du message. Les services supportés sont : la confidentialité, l'authentification, l'intégrité, la signature numérique et la non-répudiation. L'authenticité des clés publiques est vérifiée par un réseau d'utilisateurs qui certifient transitivement et avec plusieurs niveaux de confiance les clés des autres ; on n'a pas besoin d'un réseau de CA. La certification de la clé d'un utilisateur donné passe par une vérification personnelle de l'association existant entre la personne et sa clé publique. Les utilisateurs PGP distribuent leurs clés publiques certifiées par des moyens divers : signature, pages Web personnelles, etc. Le PGP utilise la cryptographie à clé publique uniquement pour protéger des clés symétriques qui sont incluses dans le message. Par ailleurs, la technologie PGP peut être utilisée pour le chiffrement des disques durs.

2.5.3 Le protocole SSH

Le but du protocole SSH (*Secure SHell*) est de sécuriser, entre autres, les commandes distantes : *rlogin*, *rsh*, *rcp*, etc. Les problèmes inhérents à l'utilisation des commandes distantes classiques sont : la faible authentification avec des mots de passe qui circulent en clair et l'authentification par l'adresse du client, ce qui implique un risque de falsification de l'adresse IP. De plus, tous les échanges subséquents (résultats issus des commandes, fichiers, etc.) circulent en clair. Avec SSH, lors de l'authentification, une clé de session est générée par le client et envoyée chiffrée avec la clé publique du serveur. Cette clé servira à chiffrer tous les échanges ultérieurs.

[19]Il a d'ailleurs risqué la prison pour cela (lire par exemple `www.absy.com/ABSMMI/ITV/ZIMM/ukitvpzim.html`).

2.5.4 Le protocole IPsec

Le modèle TCP/IP (*Transport Control Protocol / Internet Protocol*) [134] comprend quatre couches, à savoir : la couche *d'accès réseau*, la couche *Internet* (IP), la couche *transport* (TCP) et la couche *application*.

Le protocole IPsec (*IP security protocol*) est un prolongement de la famille de protocoles IP (versions 4 et 6) [134, 173]. Il fournit un certain nombre de services de sécurité et une protection de base au trafic IP avec l'authentification, l'intégrité, le contrôle d'accès et la confidentialité. Le protocole IPsec fournit les services semblables à ceux de SSL (voir plus haut), mais au niveau de la couche réseau, de sorte qu'il est complètement transparent aux différentes applications ; les applications n'ont nul besoin de prendre connaissance du protocole IPsec avant de l'utiliser. Ce protocole est d'une grande importance car la plupart des transactions commerciales électroniques sont faites à travers le réseau Internet qui utilise, entre autres, le protocole IP. Ceci veut dire que la sécurité du protocole IP est aussi importante que celle des applications installées dans les couches supérieures.

Le protocole IPsec se heurte toutefois à quelques difficultés telle que l'ambiguïté entre les utilisateurs et les adresses. Par exemple, se servir d'une adresse IP ne signifie pas qu'on en est le propriétaire. Le protocole IPsec est mieux adapté à protéger des segments Internet [99] qu'à offrir des services de sécurité aux applications et utilisateurs finaux que sont par exemple les entités du commerce électronique.

2.5.5 Synthèse : protocoles et couches du modèle TCP/IP

La figure 2.8 met en évidence les relations entre la structure en couche du protocole TCP/IP et les protocoles de sécurité. Cette figure donne une idée de l'endroit [86, 99] où il faut placer les différents services de sécurité. Dans la figure 2.8, nous avons aussi fait ressortir les sept couches du modèle OSI, créé par l'ISO, dans le but d'offrir une description standard de tout réseau informatique.

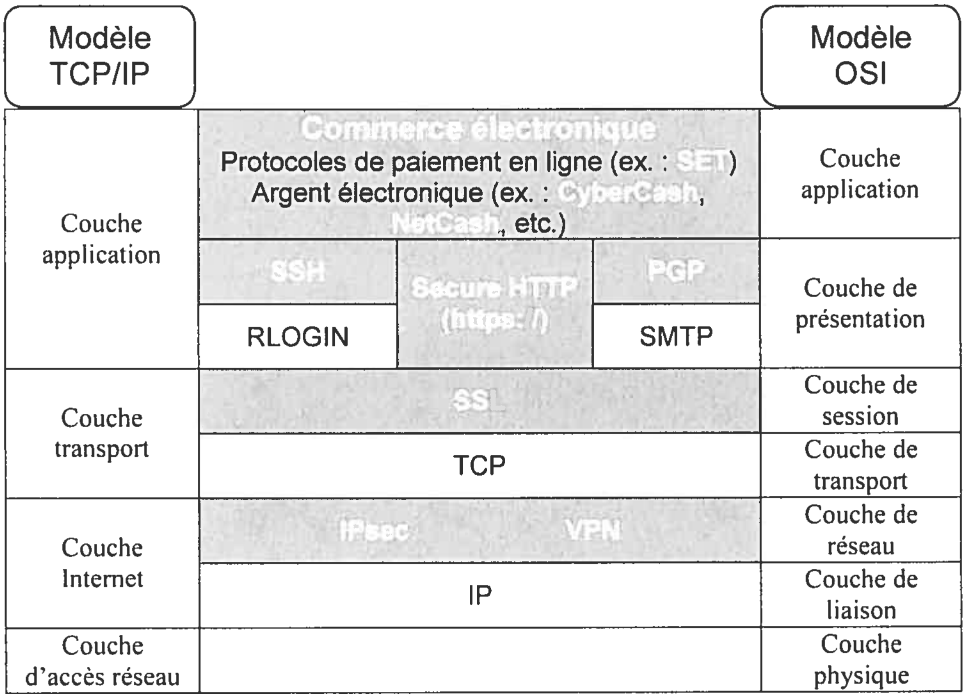


FIG. 2.8 – Couches TCP/IP et sécurité (*Inspirée d'une source anonyme*).

2.6 Conclusion

Dans ce chapitre, nous avons présenté deux types de systèmes de cryptographie qui existent sur le plan classique, à savoir les systèmes symétriques ou à clé secrète et les systèmes asymétriques ou à clé publique. Nous avons également parlé d'un certain nombre de primitives (fonctions à sens unique, fonctions de hachage, etc.) qui sont à la base de la plupart de ces systèmes cryptographiques. On se sert en fait de la cryptographie pour mettre sur pied des services de sécurité : confidentialité, intégrité, authentification, etc. C'est pourquoi nous avons jugé bon de parler des attaques, des risques et des méthodes de protection des informations échangées ou stockées. Nous reconnaissons n'avoir pas insisté sur le détail concernant les systèmes cryptographiques de renom : DES, AES, RSA, etc. En effet, de nombreux auteurs parlent de ces systèmes de long en large (description, avantages, inconvénients, attaques, etc.) et les nombreuses références que nous avons données devraient aider le lecteur en cas de besoin. Nous avons classé les systèmes cryptographiques en trois groupes selon qu'ils garantissent une sécurité inconditionnelle, calculatoire ou "démontrable". De cette classification, nous avons retenu que le seul système cryptographique qui fournit la sécurité inconditionnelle est le *masque jetable* et la grande majorité des systèmes cryptographiques restants ne garantissent qu'une sécurité calculatoire au mieux. Malgré les prouesses du "masque jetable", son utilisation n'est pas pratique car il faut une clé de longueur au moins égale à celle du message et il ne faut pas utiliser la même clé pour des messages différents, car cela donnerait de l'information à l'espion. C'est pourquoi, dans la pratique, on opte pour l'utilisation de nombreux protocoles de la catégorie "sécurité calculatoire" qui ont des longueurs de clés raisonnables et pour lesquels les clés peuvent être réutilisées. Toutefois, avec le protocole quantique de distribution de clés et les implémentations qui ont suivi (ex. : id Quantique [203]), l'utilisation du masque jetable est désormais à la portée de plusieurs modèles de commerce électronique, en l'occurrence, le B2B.

Dans le cadre du commerce électronique, nous avons noté que de nombreux protocoles y sont intégrés pour assurer les bases de sécurité requises. Bien plus, les tiers de confiance ont vu le jour pour essayer d'arbitrer les échanges entre participants des diverses transactions électroniques. Mais, ces tiers de confiance peuvent être aisément compromis et le système tout entier peut s'en trouver paralysé. La notion de calcul multi-partie a alors été présentée, son objectif étant la "suppression" des tiers à qui les utilisateurs doivent faire "confiance". Il faut toutefois noter que l'intégration des entités telles que les autorités de certification dans les processus de commerce électronique, font de celui-ci un éternel utilisateur des tiers de confiance.

Par ailleurs, nous avons présenté de nombreux protocoles dédiés au Web et qui sont utilisés dans le commerce électronique pour sécuriser les transactions, les communications, etc. À présent, nous allons mettre le cap sur la problématique de notre thèse.

CHAPITRE 3

PROBLÉMATIQUE DE VIE PRIVÉE

The privacy you're concerned with is largely an illusion. All you have to give up is your illusions, not any of your privacy
L. Ellison

3.1 Introduction

Dans le contexte du commerce électronique, le modèle CBB (Customer Buying Behaviour) [91] présenté au chapitre 1 (section 1.5) ne se soucie pas de la vie privée du client. En effet, ce modèle laisse libre champ au vendeur de colliger des données personnelles sur les clients pour en faire des profils utilisateurs. Le profil est un *dossier* qui décrit le portrait du client en termes de données démographiques (âge, sexe, nationalité, groupe ethnique, état civil, nombre d'enfants, résidence, revenu, éducation, goût, intérêts, passe-temps, etc.), tout comme il indique ses comportements d'achats et ses habitudes de navigation sur Internet. De nos jours, le vendeur dispose de plusieurs technologies [132, 173] pour constituer de tels dossiers (section 3.6).

Supposons par exemple que Sir Bob soit une célébrité du showbiz et qu'il souhaite offrir une bague de fiançaille à la belle Claudia. Supposons aussi qu'il veut faire l'achat de la bague par Internet. La bague coûte un million de dollars et Sir Bob ne voudrait pas voir son nom faire La Une des journaux à potins ! En fait, malgré son statut de star, il souhaite préserver sa vie privée autant que possible ; en particulier, il pense que sa relation avec Claudia est une affaire privée et ne devrait souffrir d'aucune forme d'indiscrétion, surtout pas celle du vendeur, Alice. Dans la suite de cette thèse, nous nous servirons de cette exemple, sous l'appellation *Showbiz*, pour illustrer les risques de violation de vie privée ainsi que les méthodes de protection que nous proposons au chapitre 5.

Comment faire pour que Sir Bob atteigne son objectif d'acheter la bague sans révéler à Alice de l'information qui pourrait plus tard porter atteinte à sa vie privée ? Voilà une question pleine de défis quand on sait toutes le étapes à franchir pour conduire d'un bout à l'autre une transaction d'achat. En fait, pour que Sir Bob atteigne son objectif, il devra être en mesure de chercher la bague dans le catalogue d'Alice, négocier le prix de vente final, payer le montant convenu, recevoir la livraison et bénéficier du service après vente sans jamais donner la possibilité à Alice de découvrir son identité ou créer un profil sur lui. En commerce électronique, pour des raisons de personnalisation, entre autres, les vendeurs vont se servir de nombreuses technologies (section 3.6) pour créer des dossiers sur les clients. Malheureusement, ces technologies vont plutôt en général constituer un moyen permettant de garder la trace des clients et de récupérer des données personnelles sur eux, pour une toute autre finalité : partage, échange, vente, etc.

Ce chapitre fait le tour du problème de vie privée en commerce électronique. Après avoir présenté le contexte de ce problème (section 3.2), nous allons tour à tour nous poser les questions QUOI, QUAND, COMMENT, POURQUOI (sections 3.4, 3.5, 3.6 et 3.7) dans le but de mieux décrire les contours du problème de vie privée en commerce électronique. À partir de cette description, en section 3.8, nous allons dégager la problématique et la motivation du travail fait dans le cadre de la présente thèse. En section 3.9, il sera question de présenter les mécanismes et solutions jusqu'ici mis sur pied en vue de protéger la vie privée en commerce électronique. Après quoi, nous allons conclure le chapitre en section 3.10.

3.2 Contexte

Depuis l'avènement de la communication orale et écrite, l'individu est contraint de faire une distinction entre les données qu'il juge publiques et celles qu'il considère comme étant privées. Si les premières peuvent être communiquées sans lui nuire, les secondes constituent plutôt des secrets à garder pour soi ou à partager avec quelques individus et organisations en qui il a *confiance*. Cette deuxième catégorie définit donc la sphère privée des données sur un individu, c'est-à-dire des données qui lui sont *personnelles*. La sphère privée existe depuis des siècles, mais sa définition n'a cessé de changer pour s'adapter progressivement à l'évolution du monde.

3.2.1 Historique et définition

Au 18^{e} siècle, plus précisément en 1763, William Pitt [43], parlementaire anglais, déclara : "*The poorest man may in his cottage bid defiance to all the forces of the Crown. It may be frail ; its roof may shake ; the wind may blow through it, the storm may enter, the rain may enter, but the King of England cannot enter ; all his force dares not cross the threshold of the ruined tenement !*" Quoi que délivré dans le cadre d'une opposition à l'instauration d'une taxe sur le cidre, ce discours met en relief l'inaliénabilité du droit à la vie privée, quelle que soit la condition humaine ! Ce discours montre aussi que les préoccupations relatives à la vie privée faisaient déjà un grand écho depuis des siècles.

En 1888, pendant la révolution industrielle, le juge Thomas Cooley [66] définissait le droit à la vie privée comme étant "*the right to be left alone*". Depuis lors, cette définition a évolué pour tenir compte des changements dans le monde, notamment ceux relatifs aux nouvelles technologies. En 1967, Westin [182] a défini la vie privée comme étant le besoin pour chaque individu de déterminer quelles données sur lui sont connues des autres individus, tout comme quand et comment ces données sont utilisées. Cette définition reste très actuelle et le Commissariat à la vie privée du Canada [190] l'a en quelque sorte adoptée en considérant que la vie privée est "le droit de contrôler l'accès à sa personne et aux renseignements qui la concernent". En d'autres termes, ces deux définitions veulent dire que chaque individu doit être maître de sa propre personne et des données qu'il juge privées. Il est donc important de veiller à ce que les données considérées comme privées par un individu ne soient utilisées ou partagées qu'avec son approbation. Par exemple, supposons que dans le contrat qu'il signe avec la banque, un client, Thomas, ne donne pas le droit ou la permission à cette dernière de partager son solde avec d'autres individus ou organisations. Si la banque annonce à un tiers qu'il y a un solde de 2$ pour l'un des comptes qu'elle administre, cette information ne causera pas de préjudice à Thomas. Par contre, si la banque dit au tiers quel est le solde de Thomas, alors cela constitue inéluctablement une violation de la vie privée de Thomas par la banque.

De nos jours, la vie privée est devenue très préoccupante sur Internet et dans un contexte de commerce électronique où Turban *et al.* [176] la définissent comme étant

"*The right to be left alone and the right to be free of unreasonable personal intrusions*". En commerce électronique justement, deux points sont importants pour ce qui a trait à la vie privée : d'une part, il faut pouvoir *classifier* une donnée comme étant privée ou publique. En effet, la notion de vie privée est relative en ce sens que ce qui est privé pour un client A ne l'est pas nécessairement pour un client B. Ceci a été illustré par Flinn et Lumsden [82], dans une étude qu'ils ont conduite en ligne pour évaluer le degré de conscientisation des utilisateurs et leurs connaissances vis-à-vis des technologies qui pourraient nuire à leur vie privée. D'autre part, il faut se demander comment faire pour *préserver* la vie privée des individus, compte tenu de la classification qu'ils ont faite de leurs données. Si la classification peut—doit— être faite par le client lui-même, les choses ne sont pas évidentes pour ce qui est de la préservation. En effet, la préservation relève non seulement du client, mais aussi et surtout des individus et organisations chargés de manipuler des données à caractère privé, dans des environnements électroniques ouverts tel que l'Internet.

Au niveau des organismes internationaux et des états, l'ONU (Organisation des Nations Unies) a institué en 1948 la *Déclaration Universelle des Droits de l'Homme* [210], dont l'article 12 est consacré à la protection de la vie privée des individus (section 3.8). Dans le même sens, une Convention Européenne sur les Droits de l'Homme a été adoptée par le Conseil de l'Europe en 1950 et est entrée en vigueur en 1953. Son article 8 met en relief, pour un individu donné, le "droit au respect de sa vie privée et familiale, de son domicile et de sa correspondance". Il a toutefois fallu attendre 1970 pour voir surgir la toute première loi au monde sur la protection des données, adoptée dans l'État du Hesse, en Allemagne. En 1973, la Suisse va suivre les traces de l'Allemagne avec le tout premier statut national sur la protection des données [109].

L'année 1980 a quant à elle connu l'adoption des lignes directrices de l'Organisation de Coopération et de Développement Économiques (OCDE) sur la protection de la vie privée et les flux transfrontiers de données personnelles [127]. Après cette adoption, une déclaration sur les flux transfrontiers de données a suivi en 1985, ainsi que celle sur la protection de la vie privée sur les réseaux, qui a eu lieu en 1988. Ces lignes directrices demeurent des instruments utiles aux gouvernements, compagnies et représentants des consommateurs dans l'élaboration des politiques de protection de la vie privée au niveau international.

3.3 Sur la violation de la vie privée

De nos jours, les moyens de communication utilisés pour l'échange des données sont très nombreux. Pour simplifier la tâche aux utilisateurs, l'Internet devient progressivement l'infrastructure commune à tous ces moyens (figure 3.1).

En accédant à Internet, l'utilisateur fait face à des menaces visibles et invisibles, des cookies au Web bugs, en passant par des Spywares, entre autres, il doit lutter sans cesse pour préserver ce qui lui est très cher : sa vie privée, donc son intimité. Cette vie privée que refusent de respecter bon nombre de vendeurs avec qui il fait affaire. Les vendeurs avancent alors plusieurs raisons. Ils décrient par exemple les nombreuses failles sur Internet : les attaques de denis de service, la fraude, le blanchiment d'argent, etc. Dès lors, ils pensent bon de mettre sur pied des mécanismes de sécurité de grande envergure, en l'occurrence, ils optent pour un filtrage à la source, une traçabilité permanente au niveau des fournisseurs d'accès Internet et des serveurs de sites Web. Par exemple, après les attentats terroristes du 11 septembre 2001 aux États-Unis, ainsi que ceux de Mardrid et de Londres, l'Union Européenne a adopté en décembre 2005 une directive qui *impose aux fournisseurs d'accès Internet et aux compagnies de téléphone de garder des données sur chaque message électronique envoyé et chaque coup de fil passé pendant une période allant de six mois à deux ans.* Toutes choses qui ont pour conséquence de remettre le respect de la vie privée à plus tard.

En réalité, les raisons qui poussent les vendeurs à porter atteinte à la vie privée de leurs clients vont au-délà du simple besoin de contrer des failles comme nous venons de le dire. En effet, l'échange, le partage et la commercialisation des données sur les clients constituent une autre raison qui explique le comportement sans scrupule de certains vendeurs qui n'hésitent pas à faire des données concernant leurs clients ce qu'ils veulent. Dès lors, même s'il faut donner raison aux vendeurs (honnêtes) compte tenu de la croissance des menaces qui les guettent, il serait utile de trouver une sorte de compro-

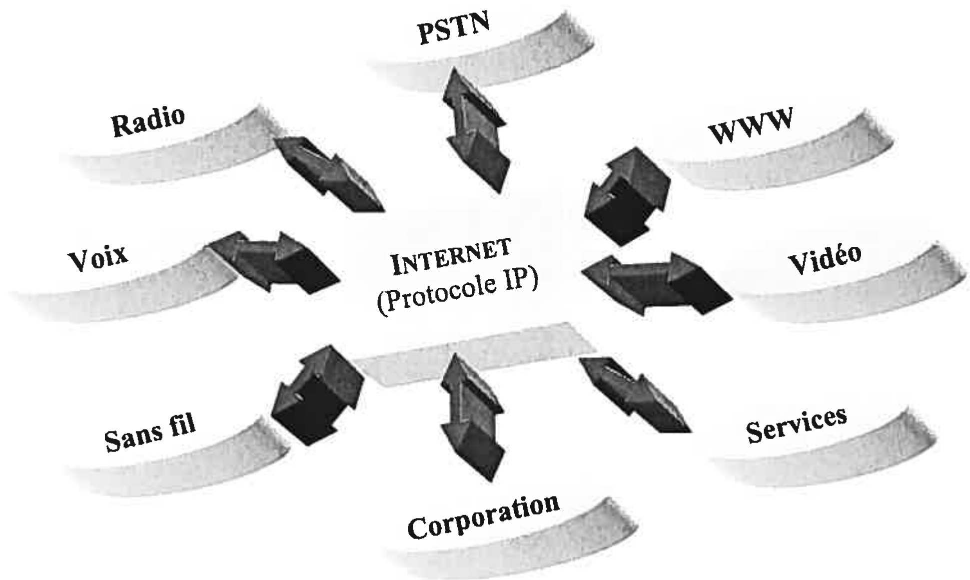


FIG. 3.1 – IP : infrastructure commune (PSTN : *Public Switched Telephone Network*).

mis entre le besoin de protéger leurs données sensibles et celui de préserver la vie privée des clients, à travers la protection des données généralement sollicités par le vendeur.

3.4 QUOI ?

La question QUOI est relative aux données personnelles des clients du commerce électronique. Ces données peuvent être regroupées en quatre principales catégories :

- **Données démographiques :** cette catégorie fait appel aux éléments suivants : âge, sexe, status matrimonial, revenu annuel, niveau d'éducation, adresses postales, co-ordonnées géographiques, numéros de téléphone, adresses de courriel, adresses Web (URL), adresses IP, entrées d'annuaires, quotient intellectuel, etc.
- **Habitudes d'achat :** il s'agit du comportement du client lors des ses achats précédents. L'intérêt peut ainsi porter sur ce que le client achète (ex. : bague, cravate en soie, chaussures en peau de serpent, etc.), l'endroit où il fait ses achats (ex. : `Metro.ca`, `eBay.com`, etc.), la fréquence d'achats (ex. : une fois par se-

maine), la fréquence de retour des produits aux vendeurs (ex. : retour de 50% de produits), les montants des dépenses (ex. : 2000$ en octobre 2005), etc.

- **Habitudes de navigation :** on analyse le comportement de navigation du client sur Internet afin de déterminer, entre autres, s'il est sensible à la publicité interactive, à certains types d'offres de produits, ou encore s'il laisse passer les cookies et autres logiciels tiers qui s'installent progressivement sur sa machine, etc.
- **Données contextuelles :** il s'agit essentiellement de l'information sur la vie du client au quotidien. Voici quelques questions d'intérêt pour ce volet : Comment le client gère-t-il ses journées ? De quels appareils dispose-t-il chez lui, au travail ou en déplacement ? Où se trouve-t-il présentement[1] : chez lui, au travail, en réunion, en déplacement ou dans une chambre d'hôtel ? Quel est son cercle relationnel ? etc.

Les preuves d'identité constituent une autre dimension des données sur le client. Il s'agit notamment [173] des preuves biométriques (ex. : empreintes digitales et rétinales, race, sexe, taille, poids), financières (ex. : numéros de comptes bancaires, numéros de cartes de crédit), légales (ex. : numéro d'assurance sociale, permis de conduire, certificat de naissance, passeport), sociales (religion, race, tribu, ethnie) et politiques (opinions politiques, parti politique). Il faut noter que les preuves d'identité et les données démographiques ont plusieurs éléments en commun.

3.5 QUAND ?

À quel moment les données personnelles sur les clients sont-elles colligées par le vendeur ? Pour répondre à cette question, il faut prendre en compte tous les moyens que le client utilise pour communiquer avec le vendeur ; c'est au moment même où le client se sert de ces moyens que le vendeur ou son partenaire d'affaire (ex. : une compagnie spécialisée dans la fabrication des fenêtres pop-ups sur Internet) va alimenter son fichier clients. Ceci inclut donc les coups de fil que le client reçoit chez lui, sa sensibilité aux

[1] La réponse à cette question est évidente quand on sait, par exemple, que le client a acheté un billet d'avion.

spams, la mise en ligne de sa page Web personnelle, l'incription dans des groupes de discussion, etc.

3.6 COMMENT ?

Les données sur le client peuvent être colligées explicitement ou implicitement, au moyen de nombreuses technologies utilisées par les sites Web des vendeurs. De telles technologies dépendent du type de données que le vendeur souhaiterait obtenir. Par exemple, pour obtenir des données démographiques ou celles sur les achats, le vendeur pourrait se servir des formulaires de saisie, alors que pour recueillir l'information sur le comportement de navigation, il pourrait utiliser, par exemple, des *cookies*, des *Web bugs*, des *spywares*, des *log files*, etc.

La réponse à la question "COMMENT ?" exige donc d'une part la présentation de quelques-unes de ces technologies et, d'autre part, les traitements que subissent les données une fois qu'elles sont à la disposition du vendeur.

3.6.1 Les cookies

Les *cookies* [193] ont été inventés par *Lou Montulli* [205] alors qu'il travaillait chez Netscape Communications. En informatique, un cookie est un petit nombre d'informations, sous forme de fichier texte, envoyées par un site Web (ex. : `Amazon.com`) à un navigateur Web (ex. : Internet Explorer, Mozilla, etc.), qui est ensuite automatiquement renvoyé lors de chaque nouvelle connexion à ce site Web. Le cookie a été inventé pour certaines données spécifiques tels que le *nom usager*, le *mot de passe*, etc. En particulier, il permet de garder les préférences du client pour des besoins de personnalisation des pages Web.

Montulli avait sûrement un objectif noble quand il pensa à inventer le cookie. En effet, étant donné que le cookie contient certaines données spécifiques, il peut aider à réduire le temps d'accès au site Web qui en est à l'origine. Toutefois, on ne saurait se limiter à cette vision de nos jours puisque le cookie est désormais en mesure de favoriser une surveillance accrue des activités d'un client sur Internet. Par exemple, il permet de stocker sur le disque dur de l'ordinateur du client les adresses des sites ou des pages Web que celui-ci a visités. Comme les cookies peuvent être lus par toute personne qui a accès à l'ordinateur du client, un simple coup d'oeil pourrait alors révéler à des tiers les sites Web visités par le client.

La bonne nouvelle est que le client peut toujours empêcher l'installation de cookies dans son ordinateur, en adjustant convenablement les options Internet de son navigateur. Il peut aussi les supprimer facilement.

La mauvaise nouvelle est qu'en bloquant les cookies, le client ne bénéficie généralement plus de certaines fonctionnalités du site Web qui souhaite s'en servir. Certains sites Web vont même jusqu'à exiger que leurs clients acceptent les cookies, au risque de ne pouvoir conduire leurs transactions électroniques jusqu'à terme.

Les cookies peuvent être installés sur la machine du client par deux moyens : soit le site Web se sert de la connection du client à son serveur via le protocole HTTP (*HyperText Transfer Protocol*), soit alors il se sert d'un langage de programmation à scripts (ex. : PHP, JavaScript, etc.) qui s'exécute dans son serveur. Avec HTTP, voici un exemple de commande de création de cookie [2] :

```
Set-Cookie: B=75r32nkld; expires=Wed, 12 May 2010 20:00;
path=/; domain=.yahoo.com
```

Dans cette instruction, `Yahoo.com` veut installer un cookie de nom *B*, de valeur *75r32nkld*, avec une date d'expiration fixée au *12 mai 2010 à 20 heures*, et ce cookie sera valable sur tous les sous-domaines du domaine `Yahoo.com` (ex. : `smallbusiness.yahoo.com`). De manière générale, le cookie a la forme suivante :

```
Set-Cookie: NAME=VALUE; expires=DATE;
path=PATH; domain=DOMAIN_NAME; secure
```

Cette écriture a déjà été expliquée dans l'exemple de cookie, à l'exception du terme facultatif "secure". Ce dernier est un paramètre qui indique que la création du cookie

[2]Dans le cas d'un language de programmation, cette commande est simplement encapsulée dans le code source du programme

par le serveur doit se faire dans un environnement sécurisé ; en utilisant par exemple le HTTPS qui est la version sécurisée de HTTP.

Malgré son côté invasif, le cookie est très utile au commerce électronique, en particulier dans les systèmes de recommandation. En effet, en combinant le cookie et le *log file* (cf. section 3.6.4), le vendeur est en mesure de mieux définir le profil du client : il sait exactement comment le client se déplace dans son site Web, à quels produits il s'intéresse, les achats fréquents, etc.

3.6.2 Les Web bugs

Un *Web bug* est une petite image de la taille d'un *point*[3], avec pour tâche la surveillance des personnes qui lisent les pages Web ou les courriels. À tout moment, le Web bug va s'identifier au site Web qui l'a créé et lui donner sa position. Contrairement aux cookies que le client peut découvrir sur son disque dur, il est difficile de voir les Web bugs. De plus, ils ne sont pas détectables par les filtres anti-cookie. Face à un site Web, rien ne peut donc signaler la présence des Web bugs au client, de sorte que le client ne peut même pas imaginer qu'on garde la trace de ses actions.

On peut se rendre compte de la présence d'un Web bug en analysant le code source HTML d'un page Web et en recherchant des *balises IMG* utilisées pour les images. De manière typique, la taille et la hauteur d'un Web bug dans une balise IMG ont chacune pour valeur 1. En outre, l'image reliée à la balise IMG devrait être chargée d'un site Web différent de celui à l'origine du code source. Voici un exemple de code source contenant un Web bug :

```
<!--INSIGHTGRIT Personal Finance Natural Search TAG START-->
<script language="JavaScript">
var s;
s='src="';
s+="http"+(document.URL.indexOf('https:')==0?'s':'')
   +"://app.insightgrit.com/1/nat?id=79152388778&ref
   ="+window.document.referrer.replace(/&/g,'|@@|')
```

[3]Semblable au point (ponctuation) qu'on place à la fin d'une phrase.

```
   +"&z="+Math.floor(Math.random()*999999)+"&purl="
   +document.URL.replace(/&/g, '|@@|');
s+='"';
document.write('<IMG WIDTH="1" HEIGHT="1" '+s+'>');
</script>
<!--INSIGHTGRIT Personal Finance Natural Search TAG END-->
```

Ce Web bug encapsulé dans du code Javascript provient du site `Quicken.intuit.com`, spécialisé dans la vente de logiciels de finance. Il est relié au serveur Web de `Insightgrit.com`.

Les compagnies de publicité tels que *DoubleClick* et *Match Point*, se servent des Web bugs pour mettre sur pied une comptabilité basée sur le nombre d'utilisateurs et de zones géographiques qui ont accès à un site Web donné. Les Web bugs aident également à déterminer le type de fenêtres publicitaires ou *pop-ups* sur lesquelles l'utilisateur est sensible et ainsi favoriser une personnalisation des pop-ups à lui présenter lors de ses prochaines connexions.

Les Web bugs utilisés dans les courriels sont davantage invasifs, puisqu'ils permettent par exemple de savoir si et quand un courriel a été lu. Ils arrivent à fournir, à l'insu de l'utilisateur, l'adresse IP utilisée pour la lecture du courriel. Plus précisément, les Web bugs peuvent être utilisés dans le but de détecter les utilisateurs qui ont ouverts des spams ; l'objectif étant alors de supprimer toutes les adresses e-mail dont les "propriétaires" n'ont pas regardé le contenu du spam. En fait, l'initiateur du Web bug suppose que ces adresses de courriel n'ont simplement pas de propriétaires.

Par ailleurs, il est possible d'intégrer des Web bugs même dans un fichier Microsoft Word. L'auteur du document peut alors traquer l'utilisation de son document : Où a-t-il été lu ? Combien de fois ? etc. Ceci est rendu possible par le fait que Microsoft Word permet de lier un document Word à un fichier image stocké dans un serveur distant, et qui n'est chargé dans le document qu'à l'ouverture.

De manière générale, le Web bug va permettre d'identifier la machine de l'utilisateur, la page Web visitée, le début de la visite, le navigateur utilisé (Internet Explorer, Mozilla, etc.) et le contenu d'un cookie que le site Web propriétaire du Web bug a pris soin de placer à l'avance.

Comme dans le cas des cookies, les Web bugs présentent quelques avantages. Par exemple, les réseaux d'annonces publicitiares peuvent s'en servir pour mettre à jour les profils des individus pour ce qui a trait aux types de sites Web que ceux-ci visitent. À partir de ce profil, il est facile de déterminer quelles publicités montrer à un individu donné. En plus, les Web bugs permettent de faire des statistiques sur les visites d'un site Web donné, ainsi que celles relatives, par exemple, à l'utilisation des navigateurs dépendamment du lieu de connection à Internet.

3.6.3 Les spywares

Un spyware est un logiciel espion qui collecte des données sur les utilisateurs et les envoie à un site Web central. Les spywares sont développés spécifiquement par des compagnies qui proposent de la publicité sur Internet. Ils intègrent trois principaux mécanismes :

- **Le mécanisme d'infection** : par exemple, le logiciel grand public de partage de musique sur Internet, Kazaa [204], vient avec un spyware appelé *Cydoor*. Cette information est d'ailleurs mentionnée dans le contrat de licence présenté à l'utilisateur au moment de l'installation du logiciel Kazaa, de telle sorte que quand l'utilisateur accepte ce contrat, il donne par ricochet la permission d'installer Cydoor et d'envoyer des données personnelles à un site central. Il faut noter que Cydoor fait partie intégrale de Kazaa et donc que sa désinstallation exige celle de Kazaa [195].
- **Le mécanisme assurant la collecte d'information** : avec l'exemple de Kazaa, la collecte va consister à enregistrer tout ce que l'utilisateur recherche et/ou télécharge par l'intermédiaire du logiciel Kazaa.
- **Le mécanisme assurant la transmission à une tierce partie** : la tierce partie peut être le créateur du logiciel ou alors une autre compagnie. La transmission se fait à travers l'Internet.

Dans tous les cas, les sypwares exploitent les machines infectées pour des fins commerciales, avec par exemple, l'affichage de nombreuses fenêtres pop-ups, le vol de données sur des individus, incluant des données financières : les cartes de crédit (exemple de spyware : *Gator*), l'analyse du comportement de l'utilisateur sur Internet (ex. : *180 Solutions*), le déroutement du trafic sur des sites web de publicité (ex. : *CoolWebSearch* et *Internet Optimizer*), etc.

3.6.4 Les log files

Un *log file* est un fichier texte qui liste au niveau du serveur Web toute action entreprise par l'utilisateur (ex. : lecture d'une page HTML, consultation d'une image ou d'un objet). En analysant un log file, il est possible d'avoir une bonne idée de la zone de provenance des visiteurs, de combien de fois ils se reconnectent au serveur, ainsi que leur comportement de navigation sur le site Web. En se servant en plus des cookies et des Web bugs, les Webmasters[4] peuvent accéder à des informations bien plus détaillées sur les différents utilisateurs qui accèdent à leurs sites Web. Voici un exemple simplifié de ligne[5] appartenant à un log file :

```
132.0.0.1 _ guerdy [11/sept/2005:21:33:00 -0700]
"GET /naissance.gif HTTP/1.0" 200 5812
```

Dans cet exemple, *132.0.0.1* indique l'adresse IP du client ; '-' indique l'absence de valeur ; *guerdy* est le nom usager du client ; *[11/sept/2005 :21 :33 :00 -0700]* donne le temps de réception de la requête du client au niveau du site Web ; *"GET /naissance.gif HTTP/1.0"* donne des précisions sur la requête du client qui demande la ressource */naissance.gif* et utilise le protocole *HTTP/1.0* ; *200* et *5812* indiquent respectivement le status et la taille du contenu de la réponse du site Web.

Les log files constituent en général le dernier récipient dans lequel attérissent toutes les données issues de l'utilisation des cookies, Web bugs, spywares, etc. C'est donc eux que le vendeur va consulter pour constituer ou mettre à jour le profil du client.

[4] Administrateurs de sites Web

[5] Nous avons dû écrire l'exemple sur deux lignes pour ne pas déborder la largeur de la page.

3.6.5 Traitement des données collectées

Les données collectées sur les clients par le vendeur sont souvent soumises à des méthodes de traitement de l'information dans le but, par exemple, de les confronter à des données détenues par un autre vendeur ou alors d'en déduire de nouvelles données. Le *data mining* [31] ou *forage de données* est la méthode de traitement la plus connue. Il s'agit d'un ensemble de techniques, automatiques ou semi-automatiques, qui permettent d'extraire de l'information utile à partir de grandes quantités de données. Les données dont il est question ici peuvent provenir d'un *entrepôt de données*, aussi appelé *data-warehouse*, ou d'une autre source tel que l'Internet. Dans le cas où le data mining est appliqué sur des données issus de l'Internet, on parle généralement de *Web mining* [163]. Le Web mining permet en particulier de faire l'analyse comportementale des clients en commerce électronique : *ventes croisées*, similarités des habitudes d'achats, etc. Les ventes croisées (*cross selling*) constituent une technique marketing dans laquelle des produits complémentaires sont proposés à un client qui a acheté ou consulté un produit donné. Par exemple, l'achat d'une imprimante à jet d'encre provoque la présentation des images de cartouche d'encre et de rame de papier au client. De même, comme c'est le cas chez `Amazon.com`, le site Web commercial peut faire savoir au client que les clients ayant acheté un produit X (ex. : bague) ont aussi acheté un produit Y (ex. : robe de mariage).

De manière générale, le Web mining permet d'améliorer les performances et le confort du site Web, de mieux rentabiliser les espaces publicitaires et de personnaliser les services pour chaque client. Les principales sources d'information du Web mining sont les log files du serveur du vendeur, les bases de données sur les clients, les cookies, etc., qui permettent d'alimenter des entrepôts de données Web ou *datawebhouses*. Il existe trois approches du Web mining :

- **Content mining** : il permet de classifier les sites Web en fonction des thèmes (ex. : sites Web académiques, d'affaires, etc.). Les contenus des sites sont alors utilisés par la suite pour classifier les clients qui les visitent. Cette approche permet par exemple de quantifier, pour chaque page d'un site Web donné, les images, les zones et la densité du texte, etc. Ainsi, en combinant le content mining et le *usage mining* (voir ci-dessous), il est possible de déterminer si les pages qui contiennent plus d'images[6] sont plus visitées que celles contenant plus de texte.

- **Structure mining** : il permet d'examiner les données relatives à la structure des sites Web. Autrement dit, il s'agit de l'architecture des sites Web et surtout des liens qui existent entre différents sites Web. Par exemple, l'analyse des chemins parcourus permet de déterminer le nombre moyen de pages Web que consultent les internautes sur un laps de temps donné et ainsi adapter l'arborescence du site Web pour que les pages les plus sollicitées soient les points d'entrée du site Web.
- **Usage mining** : il donne l'information sur la navigation des clients et l'usage qu'ils font des données disponibles sur le site Web (ex. : téléchargement, transfert de liens, référencements, etc.). Les données issues du usage mining sont stockées dans des log files et améliorent l'exploitation des cookies. Cette approche permet de mesurer la performance et l'audience d'un site Web (ex. : temps consacré à une page Web, fréquence des visites, information sur le visiteur).

Dans tous les cas, plusieurs méthodes statistiques sont utilisées dans la pratique du Web mining. Par exemple, on peut se servir des techniques d'association[7] qui permettent de déterminer les comportements *a priori* d'un client en faisant des rapprochements avec ceux d'autres clients. On peut aussi se servir de la classification dans le but de relier les caractéristiques socio-démographiques d'un client à son comportement, ou alors de la segmentation afin de rechercher des groupes homogènes pour anticiper les besoins de chaque segment de clients.

3.7 POURQUOI ?

Les données colligées à travers les technologies et issues du traitement susmentionné permettent au vendeur de faire un profilage de ses clients, en fonction des quatre types de données présentés en section 3.4. Une fois le profil du client établi, le vendeur est

[6]Généralement les pages contenant plus d'images sont plus attrayantes, mais le temps pour afficher la page peut s'avérer très long.

[7]Le filtrage collaboratif présenté à la section 1.6.3.1 en est un exemple.

en mesure de lui faire un marketing personnalisé ou marketing *one-to-one* [130], dans le but de s'assurer de sa fidélité et ainsi briser la concurrence [27]. Cette technique est semblable au marketing commercial classique, mais à des différences de comportement près. En effet, dans le monde réel [52], il est possible de faire cesser l'envoi de publicités par voie postale dans un délai raisonnable, en demandant la radiation de son identité des listes d'adresses commerciales, en s'inscrivant sur une liste rouge ou encore en apposant un autocollant contre la publicité sur sa boîte aux lettres. De telles choses sont plutôt difficiles à mettre en œuvre dans un monde virtuel.

Le marketing one-to-one suppose une collecte massive des données personnelles sur le client. Si ce dernier ne veut pas fournir de telles données, trois possibilités s'offrent à lui : il peut mettre fin au processus en cours (ex. : achat en ligne) avec le vendeur, donner de fausses informations au vendeur ou alors, naviguer de manière anonyme. Mais ceci reste le côté visible de l'*iceberg*. En effet, comme le dit si bien Dinant [75], "le simple fait d'allumer un ordinateur branché sur Internet met en route plusieurs processeurs qui exécutent subrepticement des centaines de programmes sans que l'utilisateur en soit informé, ni puisse avoir le moindre contrôle sérieux sur les données qui y sont traitées. Historiquement, cette tendance va en s'accentuant et constitue une menace majeure de plus en plus sérieuse pour la protection des données personnelles".

Dans tous les cas, les données personnelles servent essentiellement à la personnalisation des services offerts ou à offrir au client. Elle servent aussi à la publicité, au partage des données, à l'échange, à la commercialisation et à la discrimination des prix.

3.7.1 Personnalisation

La personnalisation [45, 85, 170] est une sorte de commerce sur mesure. En partant des données colligées sur des clients, le vendeur est en mesure de fournir un service personnalisé et d'adapter en temps réel la réactivité du site au comportement de chaque client. En particulier, la personnalisation permet de stocker l'historique de toutes les interactions du client avec le site Web du vendeur, dans le but de mieux cerner ses comportements d'achat et de navigation.

La personnalisation est surtout utilisée dans le commerce électronique. Elle puise en quelque sorte dans l'approche traditionnelle où le vendeur propose, de la même façon, tous ses produits à tous ses nouveaux clients. Par la suite, le vendeur propose, de manière personnalisée, une gamme de produits adaptés à chacun de ses clients.

Globalement, la personnalisation se base sur trois principes :

- La sélection de certains produits par un client conduit à la proposition de nouveaux produits, compte tenu par exemple des ventes croisées et/ou groupées (section 3.6.5).
- Selon son comportement, le client peut être rattaché à une catégorie de clients et des produits associés à cette catégorie lui sont alors proposés.
- Le profil du client, établi et mis à jour suivant son parcours à l'intérieur du site commercial et même des achats faits ailleurs, permet d'ajuster en temps réel la réactivité du site commercial vis-à-vis de ce client.

3.7.2 Publicité

La personnalisation susmentionnée permet de mieux cibler les services à offrir aux clients. En particulier, la publicité constitue un moyen efficace que le vendeur utilise pour proposer des produits à ses clients, ainsi qu'à d'autres internautes. Elle va se servir des listes de diffusion, des spams, des fenêtres intempestives (pop-ups), etc., pour intéresser le client à acheter un bien ou à visiter un site Web. Les listes de diffusion contiennent généralement des adresses e-mail fournies explicitement par les clients pendant qu'ils effectuent des achats (chez `eBay.com` par exemple). Le spam est un autre moyen d'envoyer directement de la publicité à des adresses e-mail dont l'existence n'est pas avérée à l'avance. Si le spam contient un Web bug, alors le vendeur peut retrouver des adresses e-mail valides—puisque le Web bug fera savoir au vendeur que le spam a été ouvert—et les ajouter dans ses listes de diffusion. Une analyse similaire peut être faite pour les pop-ups. En effet, en associant les clics sur les pop-ups à des Web bugs, par exemple, le vendeur est en mesure de définir le type de pop-ups à envoyer à une adresse IP donnée.

3.7.3 Partage et Échange

Le vendeur peut décider de transférer sa base de données de clients à des tiers (ex. : compagnies d'assurance, partenaires d'affaires). Cette façon de faire est considérée comme étant illicite si le client n'a pas été préalablement informé de l'éventualité d'un transfert de ses données personnelles à des tiers. L'échange est un cas particulier du partage : plusieurs vendeurs se servent d'une même base de données pour stocker et se servir des données colligées auprès de leurs clients respectifs. En d'autres termes, l'échange fait appel à la centralisation de l'information et a pour conséquence la consultation par diverses entités en ligne dont les visées ne sont pas connues des différents clients.

3.7.4 Commercialisation

Un vendeur peut décider d'utiliser les données personnelles sur ses clients comme fonds de commerce et ainsi les vendre auprès d'autres sites commerciaux. En effet, le vendeur peut considérer ces données comme faisant partie du capital de son *affaire*, de sorte qu'il peut se permettre de les commercialiser si nécessaire (ex. : en cas de faillite). Si la loi a dû remettre certains sites Web commerciaux au pas pour ce qui a trait à la volonté de vendre des données sur leurs clients (ex. : Toysmart.com et Boo.com qui ont essayé de vendre leurs fichiers clients [52]), il faut toutefois admettre qu'il est difficile de toujours stopper de telles initiatives. En fait, il est clair qu'une vente de données faite *dans le noir*, c'est-à-dire sans faire de déclaration publique, sera difficile à détecter. Surtout que le vendeur pourrait simplement faire valoir un vol de données dans son système.

3.7.5 Discrimination

Le vendeur pourrait appliquer une politique de prix discriminatoire. Plus précisément, il pourrait fixer le prix du produit en fonction des caractérisques démographiques du client. Par exemple, les prix pourraient être plus élevés pour les clients jugés plus riches.

3.8 Problématique de vie privée

Dans les sections précédentes, nous avons présenté, entre autres, les types de données sollicitées par les vendeurs en commerce électronique, ainsi que les moyens qu'ils mettent en œuvre pour y parvenir. De cette présentation, il ressort que le client du commerce électronique constitue la pièce centrale d'une infrastructure technologique révolutionnaire, mais de plus en plus invasive. Une fois sur Internet, l'utilisateur subit des assauts visibles et invisibles des technologies qui veulent à tout prix en savoir davantage sur lui [177]. Des *cookies*, aux *spywares*, en passant par des *Web bugs* et autres *log files*, *phishing*, etc., le client fait face à des choix difficiles, surtout dans le cas où il compte faire des achats en ligne. En effet, en commerce électronique, le vendeur est en mesure de colliger, de manière implicite ou explicite (cf. section 3.4), des données personnelles sur les clients. Même si la personnalisation reste compréhensible à cause de son souci de mieux satisfaire le client, il n'en demeure pas moins que sa grande consommation de données personnelles et les contours des concepts de publicité, de partage, d'échange et de commercialisation des données personnelles restent difficiles à cerner. En effet, même si la loi définit un cadre de gestion des données personnelles, elle ne dispose par de métriques pour dépister et/ou quantifier la violation de vie privée. C'est à l'utilisateur de se plaindre auprès des autorités compétentes et ainsi prouver qu'il y a eu utilisation illicite de ses données personnelles. Une telle démarche ne va toutefois pas restaurer le préjudice subi. Bien au contraire, elle ne fera que l'amplifier, compte tenu du caractère public des audiences.

De manière générale, un vendeur sans scrupules pourrait constituer une source de nombreux abus vis-à-vis du client. Il pourrait partager les données personnelles du client avec d'autres vendeurs et/ou des gouvernements. Il pourrait également vendre ces données. Ceci pourrait aboutir à une violation de la vie privée du client. De telles violations sont interdites par des lois gouvernementales mais, pour deux raisons principales, ces lois ne constituent pas en soi une garantie. Comme première raison, les *moyens technologiques* qui sont à la disposition du client ne lui permettent pas de surveiller continuellement l'utilisation par le vendeur de données sur lui. En effet, tout site

Web commercial dispose d'une politique de confidentialité que le client est tenu de lire avant la validation de la transaction électronique qui le lie au vendeur. La politique de confidentialité est toutefois une sorte de déclaration sur l'honneur que le vendeur a tôt fait d'oublier une fois la transaction complétée. La deuxième raison tient simplement au fait que les gouvernements ne sauraient être à la fois *juge* et *partie*. En effet, les gouvernements ont toujours maintenu un droit de regard, sous la forme d'un *Big Brother* [56], sur les transactions faites par des individus et compagnies. Ceci constitue sans doute une violation de la vie privée des individus.

Selon une étude menée par *Ipsos Reid* [202], en mai 2005, sur les services du gouvernement du Canada, 72% des répondants considèrent qu'il est important de maintenir la sécurité et la confidentialité des renseignements personnels. L'étude montre aussi que 48% pensent que le gouvernement du Canada possède une vaste base de données regroupant tous leurs renseignements personnels et, plus important encore, 69% sont d'accord avec l'énoncé : "Le gouvernement peut obtenir toute information voulue sur quiconque". L'étude montre par ailleurs que les individus s'inquiètent chaque jour davantage de la violation de leur vie privée. En effet, 48% des répondants estiment qu'il est très probable qu'ils soient victimes d'une atteinte grave à leur vie privée dans les deux prochaines années, tandis que 55% soutiennent que la confidentialité dans leur vie est moins bien protégée qu'il y a 5 ans.

Il est donc difficile de rassurer les internautes en général et les clients du commerce électronique en particulier, qui deviennent de plus en plus sceptiques pour ce qui est de la préservation de leur vie privée. Nous croyons que la vie privée est un droit fondamental pour tous les humains et chaque moyen visant son respect devrait être considéré avec beaucoup d'intérêt. En particulier, aucun individu ne devrait jamais devoir justifier sa volonté de protéger sa vie privée et une telle volonté ne devrait jamais être considérée *a priori* suspecte. L'article 12 de la *Déclaration Universelle des Droits de l'Homme* [210] stipule : "Nul ne sera l'objet d'immixtions arbitraires dans sa vie privée, sa famille, son domicile ou sa correspondance, ni d'atteintes à son honneur et à sa réputation. Toute personne a droit à la protection de la loi contre de telles immixtions ou de telles atteintes". Nous sommes conscients que notre position idéaliste pourrait être considérée comme étant extrême par certains. Elle pourrait même aller dans le sens contraire de la loi dans certains pays. Toutefois, nous pensons que le problème de la vie privée mérite des études sérieuses pour offrir des moyens à ceux des clients qui tiennent à la préserver, même si cela a lieu occasionnellement et/ou dépendamment du site commercial. De telles études ont déjà été faites dans d'autres contextes (ex. : Chassigneux [53] s'est penchée sur le problème, mais d'un point de vue légal), tout comme de nombreuses initiatives ont été prises par les chercheurs, gouvernements, organisations et associations de consommateurs, afin de contrer les atteintes à la vie privée.

Avant de décrire, au chapitre 4, notre approche pour solutionner le problème que nous venons de décrire, nous allons d'abord présenter quelques travaux et initiatives visant protection de la la vie privée.

3.9 Protection de la vie privée en commerce électronique

De nos jours, il existe de nombreux moyens et méthodes qui ont pour objectif la protection de la vie privée des clients en commerce électronique. Nous les avons regroupés en trois principales composantes : les moyens légaux, les moyens organisationnels et les méthodes cryptographiques. Nous présentons ces trois composantes ci-dessous.

3.9.1 Moyens légaux

À ce jour, de nombreux gouvernements ont mis sur pied des lois afin d'assurer le respect de la vie privée des utilisateurs d'Internet. Nous présentons ici les lois au Canada (Québec) et en France. Bien sûr, chaque pays a sa façon de faire. Pour une couverture plus large, il faudrait par exemple consulter [53]. Nous avons juste voulu montrer que les gouvernements élaborent et font adopter des lois pour tenter de protéger les individus contre des atteintes à leur vie privée.

3.9.1.1 Le Canada et le Québec

Le Canada compte deux lois sur la protection des renseignements personnels : la *Loi sur la Protection des Renseignements Personnels* (LPRP) et la *Loi sur la Protection des*

Renseignements Personnels et les Documents Électroniques (LPRPDE).

La LPRP a été créée en 1982 et est entrée en vigueur le 1er juillet 1983. Elle impose une limitation de la collecte, l'utilisation et la communication de renseignements personnels aux ministères et organismes fédéraux pour des besoins de respect de la vie privée. Elle confère aux individus le droit d'avoir accès aux renseignements personnels les concernant, détenus par des organisations fédérales, et celui de pouvoir les corriger si nécessaire.

La LPRPDE quant à elle a été adoptée en 2000 et est entrée en vigueur depuis le 1er janvier 2001. Cette loi vise essentiellement les renseignements personnels recueillis, utilisées ou divulgués dans le cadre d'une activité commerciale. Elle s'appuie sur dix principes qui vont de (1) la responsabilité à (10) la possibilité de porter plainte, en passant par (2) la détermination des fins de la collecte des renseignements, (3) le consentement, (4) la limitation de la collecte, (5) la limitation de l'utilisation, de la communication et de la conservation, (6) l'exactitude, (7) les mesures de sécurité, (8) la transparence et (9) l'accès aux renseignements personnels. Le détail de chacun de ces principes peut être consulté dans [207]. La LPRPDE confère aux individus le droit d'avoir accès aux renseignements personnels les concernant, détenus par des organisations et de pouvoir les corriger si nécessaire.

Pour mettre de l'emphase sur l'application des deux lois susmentionnées, ainsi que la sensibilisation continue de l'opinion publique sur des questions relatives à la vie privée, le gouvernement du Canada a institué un Commissariat à la vie privée [190]. Ce Commissariat examine les plaintes déposées par les internautes et applique des lois fédérales en cas de culpabilité. Il publie des informations sur les pratiques à appliquer dans les secteurs public et privé en matière de traitement des renseignements personnels.

Au Québec, il existe deux lois supplémentaires sur la protection de la vie privée : la *Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels* (similaire à la LPRPDE) et la *Loi sur la protection des renseignements personnels dans le secteur privé*. De plus, la Commission d'accès à l'information [191] joue au niveau de la province du Québec, un rôle similaire à celui du Commissariat à la vie privée du Canada. Plus précisement, la Commission d'accès à l'information administre la loi sur l'accès aux documents des organismes publics (ministères et organismes gouvernementaux, municipalités et organismes qui en dépendent, institutions d'enseignement, établissements de santé et des services sociaux, etc.) et sur la protection des renseignements personnels. C'est aussi elle qui veille à l'application de la loi sur la protection des renseignements personnels dans le secteur privé (entreprises de biens et services).

3.9.1.2 La France

En France, la Commission Nationale de l'Informatique et des Libertés (CNIL) [192] a été instituée en janvier 1978 par la loi relative à l'informatique, aux fichiers et aux libertés. La CNIL est une autorité administrative indépendante qui a pour mission principale la protection de la vie privée et des libertés individuelles et publiques. Pour ce faire, elle informe les personnes de leurs droits et obligations et propose au gouvernement (français) les mesures législatives ou réglémentaires de nature à adapter la protection des libertés et de la vie privée à l'évolution des techniques. Par ailleurs, la CNIL vérifie l'application de la loi à travers le contrôle des applications informatiques et établit des normes simplifiées pour des traitements de données qu'elle juge plus courants mais moins dangereux pour les libertés individuelles et publiques.

3.9.2 Moyens organisationnels

Compte tenu de la situation qui prévaut sur la violation de vie privée sur Internet, de nombreuses organisations indépendantes ont vu le jour depuis quelques années. Nous présentons ici un organisme de défense de la vie privée, l'EPIC (*Electronic Privacy Information Center*), ainsi que deux standards technologiques, P3P (*The Platform for Privacy Preferences Project*) et PET (*Privacy Enhancing Technologies*).

3.9.2.1 EPIC et compagnie

EPIC est un centre de recherche d'intérêt public créé en 1994 par David Sobel et Marc Rotenberg. Son but est d'informer le public sur les questions de libertés civiles et de protéger la vie privée dans des réseaux électroniques. Pour EPIC, plus il y a d'échanges d'information entre compagnies, gouvernements et services de police, plus les pratiques visant à traquer des individus prennent de l'ampleur sur Internet et en commerce électronique. L'EPIC met donc à la disposition des internautes en général, et des clients du commerce électronique en particulier, une boîte à d'outils contenant des solutions de protection de la vie privée. Cette boîte à outils est disponible sur Internet (`www.epic.org/privacy/tools.html`). Elle contient, entre autres, des solutions pratiques pour la sécurisation des courriels, des anonymiseurs, des contrôleurs de cookies, des générateurs de politiques de confidentialité, des protecteurs de mots de passe, etc. Cependant, l'EPIC se refuse de faire du lobbying, d'être consultant ou de jouer un rôle de conseiller pour quelque compagnie que ce soit. Il ne prend pas non plus le risque d'approuver un produit ou un service spécifique. La boîte d'outils qu'il propose constitue simplement un jeu de logiciels qui sont supposés préserver la vie privée des clients. Il n'y a donc aucune garantie de succès quant à leur utilisation. Plus grave, des questions au sujet d'un outil donné doivent être adressées directement à la compagnie ou à l'individu qui en est l'auteur(e). Malgré tout, l'EPIC joue une rôle important non seulement dans la sensibilisation de l'opinion publique, mais aussi par le fait de dénoncer tout site Web commercial qui présente des insuffisances dans la gestion de la vie privée. Pour plus d'information, le lecteur peut visiter le site de l'EPIC : `www.epic.org`.

Notons pour terminer qu'il existe de nombreux organismes de consommateurs et de défense de la vie privée. On peut citer, entre autres : *Web & Sécurité*, `websec.arcady.fr` ; *Bug Brother*, `bugbrother.com` ; EFF (*Electronic Fondation Frontier*), `eff.org` ; TRUSTe, `truste.org` ; *Privacy Rights Clearinghouse*, `privacyrights.org` ; *Junkbuster*, `junkbuster.com` ; *Privacy International*, `privacyinternational.org`.

Pour ce qui est de *Privacy International*, il est intéressant de noter que chaque année, il remet un trophée, appelé *Big Brother Awards* et représenté par une tête sous une botte [212], aux gouvernements et compagnies privées qui se sont illustrés par des atteintes à la vie privée dans leurs pays respectifs. C'est une initiative qui date de 1998 en Grande-Bretagne et, à ce jour, douze autres pays l'ont adoptée [53] : Allemagne, Autriche, Belgique, Bulgarie, Danemark, Espagne, États-Unis, Finlande, France, Hongrie, Pays-Bas et Suisse. Privacy International prime aussi chaque année les compagnies et individus qui ont largement contribué à la lutte contre la violation de vie privée.

3.9.2.2 P3P

P3P est un standard industriel développé en 2000 par le W3C (World Wide Web Consortium) dans le but de mettre à la disposition des internautes un moyen automatisé qui leur permette d'avoir un meilleur contrôle de leurs données personnelles. En se servant du standard P3P, l'internaute est en mesure de comparer la politique de vie privée mise en place sur un site Web donné avec ses préférences en termes de collecte, de traitement et d'utilisation de ses données personnelles. Le résultat de cette comparaison lui permet de décider si oui ou non il peut naviguer sur le site Web dont il est question. P3P est de nature à favoriser l'établissement d'un climat de confiance entre les internautes et les sites Web qu'ils visitent. Le principal point faible de P3P est qu'il ne fixe pas de normes minima pour la vie privée et ne fournit pas non plus de mécanismes pour s'assurer que les sites Web agissent en conformité avec les politiques qu'ils font lire aux internautes. Aussi, P3P n'est pas en mesure d'imaginer tous les cas qui pourraient surgir quand une compagnie dispose de données sur un individu. Ce qui signifie que ce standard ne saurait remplacer les politiques de vie privée sur support écrit, disponibles sur les sites Web.

P3P peut aider les gouvernements et compagnies à renforcer les politiques de vie privée sur des sites Web. Même si P3P ne s'intéresse pas aux mécanismes de transfert ou de sécurité des données, il peut quand même être intégré dans des outils conçus pour de tels mécanismes.

Enfin, il existe une version compacte de P3P (*compact policies*) dont les politiques conduisent à une interprétation rapide des pratiques d'un site Web donné en rapport avec les cookies. C'est le cas, par exemple, de l'implémentaion du P3P dans *Internet Explorer 6.0* pour la gestion des cookies, avec essentiellement six niveaux d'acceptation/refus [206] : *Accepter tous les cookies*, *Basse*, *Moyenne*, *Moyenne haute*, *Haute*, *Bloquer tous les cookies*. D'autres outils ont été développés récemment pour aider les utilisateurs à exprimer les spécifications P3P qui décrivent leurs préférences en matières de vie privée. C'est le cas par exemple de *Privacy Bird* [69] qui se veut indépendant du navigateur utilisé et, par conséquent, offre un confort beaucoup plus fin que les navigateurs commerciaux (ex. : Internet Explorer dont nous venons de parler). Il existe aussi des outils, comme par exemple SPARCLE [42], qui aident les concepteurs de site Web à décrire leurs spécifications P3P, afin que les utilisateurs puissent décider, à l'aide de *Privacy Bird* ou d'un autre système, s'ils sont satisfaits de la gestion de leurs données privées. Par ailleurs, *Bugnosis* [112] est un outil qui permet d'alerter l'utilisateur d'une site Web advenant la présence d'un Web bug, mais de l'aveu même de son concepteur, *Bugnosis* souffre de ne pas tenir compte des spécifications P3P.

Pour plus de détails au sujet de P3P, il est intéressant de visiter le site `w3.org/P3P`.

3.9.2.3 PET (Privacy Enhancing Technologies)

Les internautes doivent fournir beaucoup d'efforts et consacrer un temps considérable pour maintenir leur vie privée sur Internet. PET [46] est un ensemble d'applications logicielles ou dispositifs physiques et même des documents publiquement accessibles qui peuvent aider l'internaute à préserver ne serait-ce qu'une infime partie de sa vie privée sur Internet.

Le principe général de PET est qu'il ne faut transmettre l'information qu'à ceux qui en ont besoin pour réaliser la tâche qui leur est confiée et uniquement pour le temps qu'il faut. Ceci a pour but la minimisation du trafic des données personnelles, ainsi que leur destruction qui devrait suivre après le temps d'utilisation initialement fixé. Pour PET, ce principe devrait être appliqué sur Internet tout comme dans le monde réel. Dans le cadre du commerce électronique, par exemple, les parties impliquées sont, entre autres, le client, le vendeur, les banques, le service de livraison, le fournisseur d'accès Internet, etc. En principe, le vendeur n'a pas besoin de connaître l'identité du client, mais il doit être rassuré sur le moyen de paiement utilisé. La banque du client n'a pas besoin de connaître le vendeur ou ce que le client a acheté ; elle doit juste avoir une référence du compte bancaire à créditer ainsi que le montant en jeu. La banque du vendeur, quant à elle, n'a besoin d'aucune information sur le client. La société de livraison n'a pas besoin de connaître l'identité du client ou ce qui a été acheté—tout au plus, elle peut prendre connaissance des dimensions physiques de ce qui a été acheté—mais elle doit connaître l'identité et l'adresse du destinataire[8]. En déhors des caractéristiques techniques de la connexion Internet, le fournisseur d'accès Internet ne doit pas disposer d'information sur la transaction entre le client et le vendeur. La norme ISO 15408, relative à l'évaluation des systèmes de sécurité des technologies de l'information, met en relief des critères communs à toute protection de la vie privée, sous la forme d'une classe fonctionnelle de 4 propriétés :

- **Anonymat** : c'est le fait pour un utilisateur de ne pas être identifiable étant donné un ensemble d'objets, appelé *ensemble d'anonymat* [131]. Les transactions électroniques constituent un exemple de tels objets.
- **Pseudonimité** : elle est semblable à l'anonymat à la seule différence que l'utilisateur peut être tenu responsable de ses actes.
- **Non-traçabilité** : c'est l'impossibilité pour d'autres utilisateurs d'établir un lien entre les différentes transactions faites par un même utilisateur.
- **Non-observabilité** : c'est l'impossibilité pour d'autres utilisateurs de déterminer si une transaction est en cours.

PET propose toutefois que certaines données personnelles soient fournies aux autorités judiciaires en cas de litige ou d'enquête, par exemple dans le cadre de la lutte contre le blanchiment d'argent. Par conséquent, Les solutions utilisant le principe PET favorisent la pseudonymité au lieu d'un anonymat total. Les technologies du PET ciblent cinq secteurs principaux pour protéger la vie privée : la gestion d'entités multiples, la protection des adresses IP et de la localisation, l'accès anonyme à des services, l'autorisation respectant la vie privée et la gestion des données personnelles.

Gestion d'identités multiples : elle permet de réduire les liens entre un individu et les données le concernant. Elle exige donc que l'individu fasse des échanges anonymes avec les autres individus, mais aussi qu'il accède de manière anonyme aux infrastructures des

[8]Le destinataire peut être différent du client.

autres individus. Le problème majeur ici est que l'individu peut être traqué d'une session à une autre. S'il change de pseudonyme à chaque connexion, alors il devient difficile de bénéficier d'accès personnalisé. Toutefois, pour des rôles différents, différents pseudonymes devraient être utilisés par l'individu.

Protection de l'adresse IP : L'Internet est basé sur le principe du transfert de l'information d'un ordinateur vers un autre. Comme exemple d'information, on peut citer : les données colligées sur les clients lors de leurs visites de sites Web, l'envoi et la réception de courriel, l'utilisation de groupes et de forums de discussion, etc. Pour des besoins de transfert d'information, chaque ordinateur a justement besoin d'une *identification* unique appelée adresse IP (*Internet Protocol*). L'adresse IP peut être dynamique (changeant presqu'à chaque connexion à Internet) ou statique (fixe). La plupart des utilisateurs qui disposent d'une connection Internet fixe (ex. : modem câble) ont des adresses IP fixes, tandis que les connections dites *dial-up* se voient souvent attribuées dynamiquement une adresse IP à chaque connexion. Il faut pouvoir aider les utilisateurs à ne pas être retracés à partir de leurs adresses IP. Nous y reviendrons en section 3.9.4.2.

Accès anonyme à des services : Il s'agit d'utiliser des relais d'anonymat (*anonimity proxy*) pour naviguer sur Internet, faire du FTP (*File Transfer Protocol*), lire les courriels, etc. Ceci inclut les pseudonymes utilisés dans différents sites Web ; par exemple, plusieurs adresses e-mail créées auprès de différents portails Internet (ex. : `Yahoo.com`, `Google.ca`, etc) constituent une multiplicité d'identités virtuelles. Une telle multiplicité s'oppose à des initiatives *single-sign-on* de *Microsoft Passport* qui encouragent les utilisateurs à se créer une seule identité virtuelle et à s'en servir un peu partout sur Internet. Toutefois, *Liberty Alliance* s'attelle à gérer des identités multiples sur Internet en se servant d'un seul serveur. Les identités dont il est question ici peuvent par exemple être issues des données de connexion de l'utilisateur auprès de plusieurs fournisseurs de services : `Google.com`, `Yahoo.com`, `Amazon.com`, `eBay.com`, etc. Liberty Alliance manifeste ainsi sa volonté de protéger la vie privée des utilisateurs car cette alliance soutient que seul son serveur pourrait faire des liens entre les différents comptes d'un utilisateur donné, ce que les différents fournisseurs de services ne sont pas en mesure de faire. Malheureusement, la protection de la vie privée ici repose sur un tiers de confiance : le serveur de Liberty Alliance !

Autorisation sur Internet : En commerce électronique, les transactions lient généralement plus de deux parties (contrairement à l'approche classique "client-serveur"). Ces parties ont souvent des intérêts différents, voire opposés, ce qui peut alors engendrer une suspicion mutuelle. Il faut donc repenser les preuves d'autorisation. Ces dernières se résument en des certificats numériques (cf. section 2.4.8.2).

Gestion des données personnelles : Il faut minimiser les données personnelles. Ceci suppose la fragmentation des données, ainsi que leur anonymisation et appauvrissement (ex. : remplacement du code postal par l'identification de la région). Une solution ici est l'utilisation de PIR (*Private Information Retrieval*) présenté en section 3.9.3.3. Il faut également intégrer une auto-détermination des données, c'est-à-dire que celui qui fournit des données sur lui-même doit pouvoir contraindre l'usage qui pourrait en être fait (ex. : à effacer dans 4 heures de temps). Il faut enfin donner la possibilité de négociation entre l'individu et la compagnie (ex. : coupons de réduction en échange d'une publicité ciblée).

Un exemple de système qui base son existence sur PET est le projet PRIME (*Privacy and Identity Management for Europe*) [211]. Ce projet a pour objectif le développement de solutions pour assurer la protection de la vie privée et la gestion des identités, face aux risques issus de la révolution numérique. PRIME met en œuvre les aspects légaux et socio-économiques pour atteindre cet objectif. Il applique les principes de PET du côté utilisateur, tout comme du côté site Web et du réseau de communications. Pour plus de détails, il est intéressant de visiter le site Web [211] du projet.

3.9.3 Méthodes cryptographiques

La cryptographie (chapitre 2) est un autre moyen qui peut aider le client à protéger son identité et son profil sur Internet. Dans cette section, nous présentons Chaum et sa célèbre lutte contre la création de dossiers sur les individus dans des organisations[9]. Nous décrivons aussi la notion de transfert inconscient (OT, *Oblivious Transfer*) et celle d'extraction confidentielle de l'information (PIR, *Private Information Retrieval*).

3.9.3.1 Chaum : précurseur des transactions sans identification

Dans le cadre de cette thèse, nous nous sommes grandement inspiré des travaux de Chaum, notamment sa technique de mix-nets [54] et sa légendaire lutte contre Big Brother [56]. Chaum est le précurseur de la notion de transactions sans identification. Plus précisément, il a introduit une approche qui protège les données personnelles que les organisations privées et publiques échangent le plus souvent dans le but de faire des statistiques et du data mining, de créer des profils des clients, etc. Cette approche protège aussi les compagnies contre les possibles abus des clients et, d'une manière générale, elle peut être appliquée à trois types de transactions : la communication, le paiement et les lettres de créances.

- **Communication** : proposition d'un système dans lequel l'individu donne des pseudonymes différents à chaque organisation avec laquelle il fait des affaires, de telle sorte qu'il est impossible que ces organisations puissent se concerter et constituer des dossiers sur lui.
- **Paiements** : introduction de la monnaie électronique intraçable pour permettre à un client d'acheter des produits sans laisser de traces sur ses paiements.
- **Lettres de créances** : introduction de mécanismes qui permettent à un individu de prouver à une organisation qu'il dispose des créances nécessaires à une tâche, sans donner explicitement de l'information sur l'ensemble des créances dont il dispose.

En 1981, Chaum a introduit la technique du mix-nets, basée sur le cryptographie à clé publique alors naissante (cf. section 2.4.2). Cette technique permet au système d'envoi de courrier électronique (courriel) de cacher l'identité de l'émetteur du courriel, ainsi que le contenu du message lui-même (*untraceable electronic mail*). Elle permet aussi au receveur du message de disposer d'une adresse intraçable pour le courriel de retour (*un-*

[9]Sa lutte date d'avant l'Internet grand public.

traceable return address). Cette technique demeure calculatoirement sécuritaire même si l'infrastruture de communication sur laquelle elle s'appuie est insécure. Elle a l'avantage de ne pas nécessiter un tiers de confiance *commun*. Au contraire, chaque participant dans le cheminement du courriel peut être comme une autorité, de sorte que la solution de Chaum ne peut être compromise qu'en cas de subversion ou corruption d'un nombre significatif de participants.

Dans un contexte de commerce électronique, la technique du mix-nets peut par exemple aider le client à envoyer des requêtes à un serveur Web sous la forme d'un courrier électronique intraçable et à recevoir des réponses via l'adresse de retour intraçable.

3.9.3.2 Transfert inconscient

La notion de transfert inconscient ou OT (*Oblivious Transfer*) a été à l'origine conçue par Wiesner [183] dans un papier qu'il écrivit en 1970 mais qui ne fut puplié qu'en 1983. Elle a été indépendamment réinventée dans une forme plus simple par Rabin [137] en 1981. Le scénario de Rabin est le suivant : Alice (l'émettrice) a un bit en tête et elle veut le transmettre à Sir Bob (le receveur), mais cela doit se faire à travers un canal qui a une probabilité de $\frac{1}{2}$ de le laisser passer, c'est-à-dire que Sir Bob va recevoir le bit avec une probabilité de $\frac{1}{2}$. Sir Bob saura s'il a reçu le bit ou pas, mais Alice ne le saura pas. De plus, aucune des deux parties ne peut influencer la probabilité de 50% imposée par le canal. Après le papier de Rabin, d'autres variantes d'OT virent le jour dans le monde de la cryptologie. La première variante est le *1-out-of-2 OT*, ou OT_1^2, introduit par Even, Goldreich et Lempel en 1985 [77]. Dans un contexte de OT_1^2, Alice a deux bits (b_0, b_1) en tête. Sir Bob peut choisir et obtenir un bit, b_i. À la fin du processus, Alice ne devrait rien apprendre sur i et Sir Bob ne devrait rien apprendre sur l'autre bit, b_{1-i}. De plus, Sir Bob ne devrait pas être en mesure d'obtenir une information conjointe de b_0 et b_1 mis ensemble (le OU-exclusif des deux par exemple). Dans l'ignorance totale des travaux ci-dessus[10], une autre variation d'OT fut introduite en 1986 par Brassard, Crépeau et Robert [40], sous le nom ANDOS (*All or Nothing Disclosure of Secrets*),

[10]Communication personnelle avec Gilles Brassard, l'un des auteurs.

aujourd'hui connu sous le nom OT_1^n. Le scénario de OT_1^n est le suivant : Alice a n chaînes de caractères $x_1, \cdots, x_n$ et Sir Bob veut choisir l'une d'elles, disons x_i, pour un certain $1 \leq i \leq n$. À la fin du processus, Alice ne devrait rien avoir appris au sujet de i et Sir Bob ne devrait avoir appris aucune information conjointe sur les chaînes d'Alice. Il faut noter que Crépeau [70] demontra plus tard que toutes les variantes d'OT présentées ci-dessus sont équivalentes.

3.9.3.3 Private Information Retrieval

La notion de *Private Information Retrieval* (PIR) a été introduite par Chor, Goldreich, Kushilevitz et Sudan en 1995 [60]. PIR permet à un utilisateur d'obtenir, de manière privée, le i-ème bit, x_i, d'une chaîne de n bits, $x = x_1x_2\cdots x_n \in \{0,1\}^n$, sans révéler quelque information que ce soit au serveur. Contrairement au transfert inconscient, PIR ne garantit pas le secret des bits autres que celui sollicité par l'utilisateur. Dans la conception de base de PIR, la chaîne x est considérée comme étant une base de données stockée dans un serveur. Dans ces conditions, une façon simple de réaliser PIR est d'envoyer toute la base de données à l'utilsateur qui se chargera alors de choisir le bit désiré. Ceci porte le nom de *schéma trivial* de PIR. Le schéma trivial conduit de manière évidente à un coût de communication de n bits. La principale raison d'être de PIR est donc de minimiser le nombre de bits envoyés du serveur à l'utilisateur.

Pour avoir un complexité de communication sous-linéaire, de nombreux protocoles de PIR ont été proposés [25, 48, 60, 104, 105], dans un contexte de sécurité inconditionnelle ou calculatoire. PIR inconditionnellement sécuritaire a été introduit dans [60] ; il requiert que la requête envoyée au serveur par l'utilisateur ne donne aucune information sur i au serveur, quoi qu'il arrive. Les résulats suivants ont été prouvés dans [60]. Premièrement, tout protocole de PIR utilisant une seule base de données requiert $\Omega(n)$ bits de communication ; et donc, ne peut pas faire mieux qu'un schéma trivial. Et deuxièmement, pour avoir un coût de communication sous-linéaire, les données doivent être répliquées dans plusieurs serveurs, qui sont supposés ne pas communiquer entre eux.

Dans un modèle calculatoire, d'intéressants protocoles de PIR ont été proposés par

Chor et Gilboa [63], Kushilevitz et Ostrovsky [104, 105], Cachin, Micali et Stadler [48], Chang [51] et Lipma [108]. Ces protocoles sont basés sur des problèmes difficiles tels que le problème des résidus quadratiques [104], les fonctions à sens unique avec brèche [105], etc. Ces protocoles utilisent un seul serveur de bases de données, mais ramènent la complexité de communication à un niveau sous-linéaire.

Il existe d'autres implémentations de PIR, dont une basée sur l'utilisation d'une composante matérielle proposée par Smith et Safford [165]. La composante matérielle dont il est question ici est un co-processeur sécure qui est installé à même le serveur et fonctionne comme une *boîte noire* pour recevoir, chiffrer et transmettre l'information sollicitée par l'utilisateur. En utilisant l'idée de co-processeur, Asonov et Freytag [20] ont mis au point un protocole PIR avec un coût constant, $O(1)$, en communication et en temps de calcul. Leur protocole fait des mises à jour périodiques off-line sans nécessiter un quelconque coût en terme de communication.

Les protocoles PIR garantissent uniquement la vie privée des utilisateurs, sans songer à protéger les données du serveur. En fait, PIR permet à l'utilisateur d'obtenir des bits en plus de celui qu'il a sollicité. Pour remédier à cela, une version multiserveur de PIR symétrique ou SPIR (*Symmetrically Private Information Retrieval*) a été introduite dans un modèle calculatoire [104], d'une part et dans un modèle basé sur la théorie de l'information [87], d'autre part. Un protocole SPIR protège la vie privée de l'utilisateur tout en le restreignant à ne recevoir que l'unique bit qu'il a demandé dans sa requête (adressée aux différents serveurs). Dans le cas d'un seul serveur de bases de données, de nombreuses implémentations de SPIR ont aussi été proposées dans un modèle calculatoire [20, 87, 122].

Pour finir, il faut noter que SPIR avec une base de données de longueur n et OT_1^n sont équivalents. Toutefois, Malkin [114] note une différence entre SPIR et OT_1^n dans leurs motivations respectives : un protocole SPIR vise essentiellement à réduire le coût de communication tandis que l'efficacité du calcul est la préoccupation essentielle des protocoles OT_1^n.

3.9.4 Services d'un tiers de confiance

Pour protéger leur vie privée, les clients ont parfois aussi recours à des tiers de confiance. Nous présentons ici les tiers de confiance dans la cadre de la navigation (incluant les transactions) et de la livraion anonymes.

3.9.4.1 Nagivation anonyme

Les utilisateurs peuvent être retracés à partir de leurs adresses IP [208]. En fait, l'adresse IP est information d'identification qui est automatiquement capturée par un autre ordinateur chaque fois qu'un lien de communication est établi sur Internet. Différents moyens permettent de garder les traces d'un utilisateur.

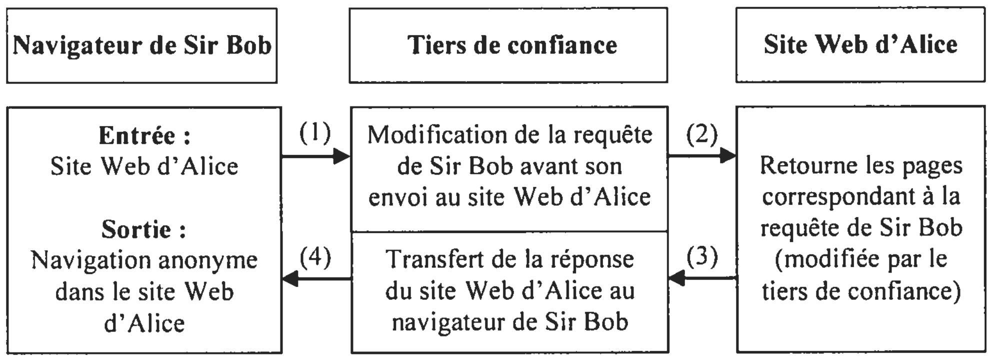
1. Les adresses IP sont distribuées par blocs aux fournisseurs d'accès Internet et des bases de données de ces blocs sont publiquement connues et accessibles pour des besoins de consultation.
2. À partir de l'adresse IP d'un ordinateur, son adresse machine peut être retrouvée en utilisant le RARP (*Reverse Address Resolution Protocol*) [134].
3. La commande *traceroute* permet de trouver le chemin suivi par les paquets d'information, de l'origine à la destination finale. Les paquets qui circulent sur Internet passent à travers plusieurs ordinateurs, dans un ordre hiérarchique. Par exemple, les paquets d'information peuvent aller d'un ordinateur de départ au Fournisseur d'Accès Internet (FAI) d'attache, jusqu'à ce qu'il atteigne l'ordinateur qui constitue le sommet de la hiérarchie ou *Backbone*. Par la suite, les paquets d'information vont être transférés du Backbone jusqu'à l'ordinateur de destination. Il se peut que les paquets n'atteignent pas le Backbone si le FAI de l'ordinateur destinataire se trouve dans la hiérarchie, entre le FAI d'origine et le Backbone.
4. Il est facile d'avoir de l'information sur les enregistrements de nom de domaines Internet, en utilisant par exemple une simple commande Unix/Linux *whois* (ex. : *whois amazon.com* donne l'information d'enregistrement du site Web `amazon.com`).

5. Il est possible de retrouver l'adresse IP et/ou le nom de l'ordinateur de d'utilisateur d'une page Web donnée. On peut par exemple se servir des technologies Web bugs et Spywares présentées en section 3.6.

Dans le cadre de cette thèse, c'est ce dernier moyen (5) qui nous interpelle le plus. En effet, si le vendeur connaît l'adresse IP ou le nom de l'ordinateur du client, alors il n'y a plus de vie privée pour ce client. Il est donc important d'aider le client à naviguer sur Internet de manière anonyme, de sorte que le vendeur soit en mesure de lui délivrer les produits numériques sans pouvoir l'identifier à travers son adresse IP, son nom, etc.

Comment peut-on alors naviguer de manière anonyme ? Compte tenu de ce qui précède, il est clair qu'il faut aider les utilisateurs à cacher les adresses IP de leurs ordinateurs dès lors qu'ils sont sur Internet. Une solution simple consiste à se servir des cafés Internet, mais ceci n'est pas toujours une solution aisée pour l'utilisateur qui doit chaque fois se déplacer, en plus de l'indiscrétion suscitée chez les autres utilisateurs du café. L'utilisation des sites Web proxy offre un certain degré de sécurité. Si le proxy est placé entre l'ordinateur et l'Internet, alors tous les utilisateurs semblent provenir d'un même ordinateur, de sorte les utilisateurs ne peuvent pas être retracés au-délà du proxy, à moins de disposer d'une information additionnelle. Comme exemple d'information additionnelle, l'on peut citer le nom de la plus grande ville située près de l'utilisateur, généralement inclus dans les noms des ordinateurs des FAI. De plus, les données d'enregistrement fournies par l'utilisateur au FAI sont soumises aux mêmes procédures de collecte et de traitement de données personnelles présentées en section 3.6. Dans tous les cas, même si l'utilisateur se sert d'un proxy, il est clair que l'emploi des Web bugs et spywares par le site Web ouvrent une nouvelle voie qui attribue un identifiant unique à l'utilisateur.

Pour résoudre le problème de violation de la vie privée lié à l'adresse IP, des *routeurs de confiance* ont vu le jour depuis 2002. Ces routeurs offrent la possibilité de naviguer sur Internet sans envoyer de données personnelles à d'autres sites Web. Ceci se fait en modifiant la requête de l'utilisateur à travers le navigateur, et en nettoyant les pages qui sont retournées à l'utilisateur. Ceci est illustré dans la figure 3.2, inspirée de [199] (dans la rubrique *About US*). Les routeurs de confiance permettent de naviguer de manière



(1) La requête de Sir Bob est chiffrée (avec la clé publique du tiers de confiance et éventuellement celle d'Alice) et envoyée au tiers de confiance
(2) Le tiers de confiance modifie la requête en cachant particulièrement l'adresse IP de Sir Bob
(3) Le site Web d'Alice retourne les pages correspondant à la requête envoyée en (2)
(4) Le tiers de confiance transfère les pages reçues en (3) à Sir Bob

FIG. 3.2 – Routeur de confiance pour cacher les adresses IP.

anonyme, en cachant l'identité de l'utilisateur ainsi que son adresse IP, et en bloquant les cookies, les fenêtres publicitaires intempestives, etc. Toutefois, cette solution suppose que le routeur de confiance est lui-même digne de confiance, et la vie privée de l'utilisateur dépend crucialement de cette hypothèse.

Si on ne veut pas faire confiance à un routeur de confiance, alors on peut se servir des routeurs d'anonymat tels que le *mix-net*, l'*Onion Routing* et les *Crowds*. La technique du mix-net a été mentionnée en section 3.9.3.1.

L'Onion Routing consiste en un ensemble de projets visant la recherche, l'analyse, la conception et l'implantation des systèmes de communications anonymes. L'accent est surtout mis sur les systèmes qui peuvent résister à l'analyse de trafic, l'espionage, etc. L'Onion Routing empêche le moyen de communication utilisé de savoir *qui* parle à *qui*, tout au plus le réseau de communication est au courant qu'une communication a lieu. De plus, l'Onion Routing empêche l'espion de disposer en clair du message communiqué ; en fait, seul le destinataire final du message ou l'endroit où le message quitte l'Onion

Routing vont avoir accès au contenu (en clair) du message.

Les Crowds [138] quant à eux permettent de grouper un très grand nombre d'utilisateurs autour d'une même adresse IP de sorte que ces utilisateurs communiquent avec les sites Web en se servant de cette unique adresse IP. Reiter et Rubin [138] soutiennent qu'il est alors difficile aux serveurs des sites Web de savoir qui est à l'origine de la communication, puisqu'elle peut provenir de n'importe quel membre du Crowd. En particulier, ces serveurs devraient faire coalition avec un nombre significatif de membres du Crowd pour arriver à obtenir l'origine de la communication.

D'autres systèmes existent avec des objetifs plus ou moins identiques à ceux des systèmes susmentionnés. Par exemple, le système *Mixmaster* (`Mixmaster.sourceforge.net`) fournit les services de *remailing*, une technique permettant de se protéger contre l'analyse de trafic et, surtout, d'envoyer des courriels de manière anonyme. Le lecteur pourrait aussi consulter le système *Tor* (`Tor.eff.org`)[11].

3.9.4.2 Livraison anonyme

La livraison anonyme des produits numériques est assurée *ipso facto* de la navigation anonyme. Dans le cas des produits physiques, la situation est beaucoup plus compliquée. En effet, il semblerait que le vendeur ait besoin de l'adresse du destinataire du produit acheté afin de faire la livraison. Or, donner cette information peut conduire à une atteinte totale à la vie privée du destinataire (dans la plupart des cas, le client lui-même). Il faut donc trouver un mécanisme qui permette au client de recevoir la livraison du produit sans que le vendeur prenne connaissance de son adresse.

Une solution consiste en l'utilisation, une fois de plus, des tiers de confiance. Dans ce cas, le client donne une procuration au tiers de confiance afin qu'il récupère le produit auprès du vendeur et l'achemine jusqu'à lui. On peut citer ici les compganies *ContinentalRelay* [194] et *Executive Mail Drop Services* [198] qui offrent des adresses privées à leurs clients pour qu'ils y récupèrent des paquets qui leur sont envoyés. Ces compagnies offrent d'ailleurs des services additionnels tels que les adresses e-mail anonymes,

[11] Dernière date de consultation pour `Mixmaster.sourceforge.net` et `Tor.eff.org` : le 20 décembre 2005.

les boîtes postales, les transferts d'appels téléphoniques et de fax, les boîtes vocales, les bureaux virtuels et voire même les services de secrétariat. Dans la plupart des cas, les produits reçus par les tiers de confiance sont simplement re-adressés et expédiés aux clients légitimes.

3.9.5 Limitations

Tous les moyens et méthodes de protection de la vie privée que nous venons de décrire présentent chacun des avantages et des inconvénients. Nous mettons ici beaucoup plus l'accent sur les inconvénients car ce sont eux que nous avons exploités pour bâtir notre approche (chapitre 4).

Les moyens légaux ont l'avantage de prendre appui sur les gouvernements qui ont la possibilité de faire respecter la loi. Toutefois, comme nous l'avons dit en section 3.8, les gouvernements sont à la fois juge et partie. Par exemple, la notion de certificat de sécurité [216] émis sur des individus accusés de terrorisme, au Canada, crée une sorte d'empiètement sur la charte canadienne des droits et libertés [189] car, une fois qu'un tel certificat est émis, l'accusé peut automatiquement être mis en détention ou faire l'objet d'une déportation vers son pays d'origine.

Les moyens organisationnels apportent toute la force qu'on connaît à la société civile. Mais ce n'est pas facile de distinguer et de valider clairement le soutien à apporter aux individus. Par exemple, comme nous l'avons dit en section 3.9.2.1, EPIC met à la disposition des individus des outils de lutte contre les atteintes à la vie privée, mais il est incapable de rassurer ces mêmes individus pour ce qui est de l'utilisation sécuritaire de ces outils. Du coup, ces outils peuvent plutôt être dangereux. Il suffit par exemple d'imaginer l'installation d'un *anti-spyware* qui est en fait un facilitateur de copie de logiciels d'espionnage sur des disques durs.

Pour ce qui est des méthodes cryptographiques, le plus grand inconvénient est qu'elles ne sont pas adaptées à gérer une transaction de commerce électronique de bout en bout. Par exemple, faire du PIR ne suffit pas à conclure une transaction d'achat qui préserve la vie privée du client. En particulier, PIR ne va pas gérer les paiements, la livraison des biens, le service après vente, etc. Bien sûr qu'il y a des travaux qui s'appuient sur les méthodes cryptographiques, pour gérer des processus d'achat/vente au complet. C'est par exemple le cas du "Priced Oblivious Transfer" de Aiello, Ishai et Reingold [7]. Toutefois, même dans ce cas ci, les auteurs vont se limiter à la vente des biens numériques.

Pour finir, en ce qui concerne les tiers de confiance, il faut donner la possibilité au client de contrôler le tiers de confiance, en le forçant par exemple à ne pas faire coalition avec le vendeur.

Nous décrivons au prochain chapitre notre approche de préservation de la vie privée des clients en commerce électronique. Contrairement aux moyens et méthodes décrits ci-dessus, notre approche gère la transaction d'achat/vente de bout en bout, de l'étape de recherche du produit à celle de service après vente, en passant par celles de négociation, de paiement et de livraison. En fait, nous sommes parti du modèle standard CBB (section 1.5) et avons développé des protocoles qui aident le client à préserver sa vie privée à chaque étape du processus d'achat. Le détail de tout ceci sera présenté dans les chapitres 4 et 5.

3.10 Conclusion

Dans ce chapitre, nous avons décrit le concept de vie privée en insistant sur les données visées par le vendeur (QUOI), le moment pendant lequel le client laisse échapper ces données (QUAND), les outils technologiques utilisés par le vendeur (COMMENT), ainsi que l'utilisation que le vendeur peut en faire (POURQUOI). Ceci nous a conduits à soulever la problématique relative à la protection des données privées des clients, compte tenu surtout de l'utilisation que le vendeur peut en faire. Nous avons par la suite présenté quelques approches de préservation de la vie privée en commerce électronique, puis passé en revue les limitations des ces approches.

La principale limitation relevée dans les méthodes cryptographiques présentées est l'absence des protocoles qui intègrent la protection des données privées à tous les niveaux du processus d'achat/vente et cela dans un contexte réaliste—pour nous, il est par exemple irréaliste de ne considérer que l'achat des produits numériques comme c'est le cas dans [7].

Notre approche s'aligne sur les rubriques de méthodes cryptographiques et de tiers de confiance. Elle tient également compte des principes énoncés par certaines entités organisationnelles, en l'occurrence, PET. Nous pensons que les moyens légaux présentent davantage de limites car, d'une part il y a ce problème d'être juge et partie et, d'autre part, la justice peut elle même constituer une source de violation de la vie privée. Par exemple, dans la cas du Showbiz, si Sir Bob a la preuve que sa vie privée a été violée par Alice, il pourrait porter plainte auprès des autorités compétentes, pour que justice soit faite. Cependant, la plainte à elle seule ne suffit pas pour rétablir sa vie privée. Au contraire, cela lui causera davantage de tort puisque sa position de star ne fera qu'intéresser encore plus de médias.

Nous avons consacré le prochain chapitre à la description de notre approche.

CHAPITRE 4

MODÈLE BCBB

Un mathématicien est un aveugle dans une chambre noire qui est à la recherche d'un chat noir qui n'est même pas là

C. Darwin

4.1 Introduction

Au chapitre 3, nous avons mis en relief la problématique liée à la violation de la vie privée en commerce électronique, ainsi que les moyens et méthodes mis en œuvre par les gouvernements, les organisations et les chercheurs en vue de la solutionner. Nous y avons également fait ressortir les faiblesses de ces moyens et méthodes. Dans ce chapitre, nous décrivons l'approche que nous avons mise de l'avant pour apporter une solution globale au problème de la vie privée en commerce électronique. Une fois de plus, notre approche est destinée non seulement à la protection des données privées du client, mais aussi à celle des informations que le vendeur juge sensibles.

La notion de vie privée en commerce électronique trouve son fondement sur l'existence d'une sphère *privée* de données sur des clients. En section 3.4, nous avons décrit les données que le vendeur pourrait solliciter du client lors d'une transaction électronique. Pour mettre sur pied une solution de lutte contre des atteintes à la vie privée, il est important que chaque client fasse une classification des ses données. Une telle classification trouve toute sa justification dans le fait que *ce qui est privé pour un client ne l'est pas forcément pour l'autre*.

Par ailleurs, nous considérons de façon similaire la sphère *sensible* des données du vendeur. En effet, la protection des intérêts du vendeur passe par une classification des données de son site Web commercial car, *ce qui est constitue une donnée sensible pour un vendeur n'en est pas nécessairement une pour l'autre*.

Notre approche consiste à transformer le modèle standard CBB (*Customer Buying*

Behaviour) en un nouveau modèle, appelé BCBB (*Blind Customer Buying Behaviour*), dans lequel le client est en mesure de fouiller dans le catalogue du vendeur, négocier le prix de vente et les conditions de vente avec le vendeur, payer, recevoir le livraison du produit et bénéficier du service après vente, tout cela sans que le vendeur apprenne quoi que ce soit sur ses données privées. Toujours dans ce nouveau modèle, s'il s'agit d'un produit numérique, le vendeur ne saura ni ce qu'il a vendu, ni le montant de la transaction[1] (Section 4.3.3). Pour ce qui est des produits physiques, le vendeur saura ce qu'il a vendu, mais encore une fois, il n'apprendra pas le montant de la transaction.

Dans la suite de ce chapitre, nous allons d'abord présenter la classification que nous avons faite des données sur le client, ainsi que de celles du vendeur (section 4.2). Par la suite, nous allons utiliser cette classification pour définir notre modèle de gestion de la vie privée du client et la sensibilité des données du vendeur. Plus précisément, nous donnerons en section 4.3 un apperçu du modèle BCBB, ainsi que celui des centres de livraison anonyme (section 4.3.2.1) qui permettent au client de recevoir la livraison de produits physiques sans compromettre sa vie privée. Par la suite, nous parlerons en section 4.3.2.2, toujours sous forme d'aperçu, de notre système de recommandation qui permet de faire des recommandations de produits aux clients tout en préservant leur vie privée. Nous poursuivrons, en section 4.4, avec la nécessité de tenir compte de la protection de l'ordinateur du client avant de conclure en section 4.5.

4.2 Classification des données

De manière générale, comme nous l'avons dit en section 3.9.2.3, une transaction en commerce électronique implique plusieurs entités, parmi lesquelles se retrouvent le client, le vendeur, les banques, le service de livraison, les fournisseurs d'accès Internet, ainsi que les intermédiaires virtuels de circonstance que sont tous les serveurs Internet par lesquels cheminent les données utiles à la transaction. Toutefois, les données de la transaction proviennent essentiellement du client et du vendeur. Elles peuvent donc être

[1]En effet, si le vendeur connaît le montant de la transaction, alors il pourrait savoir de quel produit il s'agit.

classifiées en fonction de l'importance que leur accordent ces deux entités.

4.2.1 Données du client

Les données du client dont la communication au vendeur pourraient conduire à la violation de la vie privée sont les suivantes : l'identité, les données démographiques, les habitudes d'achats et de navigation, les données contextuelles, ainsi que les besoins actuels et à venir. Ce sont là des *composantes* essentielles qu'il faut gérer en matière de vie privée. Une autre *composante* de la vie privée à prendre en compte est le *Tracking*[2] qui est le fait pour un serveur Web de garder la trace des utilisateurs au fur et à mesure qu'ils s'y connectent. Ainsi, le tracking permet au vendeur de savoir si un même client revient le plus souvent dans son site Web, même s'il ne connaît pas son identité, ses données démographiques, etc.

Nous avons pensé à séparer la *vie privée dans sa globalité* de sa composante *tracking* pour mieux préserver la vie privée des clients. En effet, bon nombre de clients trouvent que la non communication des données privées au vendeur n'est pas suffisante pour protéger leur vie privée, si le vendeur est en mesure de les traquer [172]. Cette séparation nous a conduit, **d'une part**, à la création d'une liste de contrôle, appelée SLP (*Stop List for Privacy*), formée de toutes les données décrétées privées par le client. Si une donnée de la SLP est sollicitée par le vendeur, le client réagit dépendamment du *degré de tolérance* en vigueur. Nous avons défini trois niveaux de tolérance :

1. *Tolérance zéro sur les données privées* : le client tient à garder secrètes toutes les données de sa LSP, coûte que coûte. De plus, il ne souhaite pas que le vendeur apprenne quel produit l'intéresse.
2. *Tolérance moyenne sur les données privées* : Le client tient à garder secrètes toutes les données de sa LSP, coûte que coûte, mais il peut laisser le vendeur apprendre le produit qui l'intéresse.
3. *Tolérance totale sur les données privées* : Le client n'est pas du tout gêné qu'on

[2]Ce mot ne semble pas avoir un équivalent en français. La signification évidente est "le fait d'être traqué", mais le verbe "traquer" a un sens très différent.

apprenne son identité, encore moins qu'on constitue un dossier sur lui, incluant ses données démographiques, ses habitudes d'achats et de navigation, etc.

L'utilisation de la LSP est dynamique et peut être faite suivant deux axes. Dans le *premier axe*, comme le client peut traiter avec plusieurs vendeurs, il s'en suit que sa LSP devra dépendre du vendeur auquel il fait présentement face. Par exemple, Sir Bob (Showbiz) pourrait accepter de dévoiler son identité à un autre vendeur, Éric, auprès de qui il achète une cravate, sans pour autant renoncer à sa décision d'acheter la bague de Claudia sans laisser de traces auprès d'Alice. Pour ce qui est du *deuxième axe*, on supose que le client fait affaire avec le même vendeur. Un tel client pourrait décider de faire un achat sans traces à un moment donné, puis recourrir au modèle avec traces lors du prochain achat, dépendamment du type d'achat à faire. Par exemple, Sir Bob pourrait acheter une bague et une cravate auprès d'Alice, mais en ne dévoilant son profil que dans le second cas.

D'autre part, nous avons aussi considéré trois niveaux de tracking :

1. *Tolérance zéro au tracking* : le client ne veut pas du tout laisser de traces, pour quelque raison que ce soit. Cette contrainte ne peut être respectée que si aucune entité, en l'occurrence le vendeur, n'est en mesure de savoir si un utilisateur donné est en fait un ancien utilisateur du système. Dans ce cas, s'il s'agit par exemple d'un système de recommandation basé sur le filtrage collaboratif[3] (section 1.6.3.1), le client devrait lui-même fournir toute l'information nécessaire à la production des recommandations chaque fois qu'il revient chez le même vendeur. Cette information inclut, par exemple, les votes sur les produits achetés par le passé.

2. *Tolérance moyenne au tracking* : Le client ne veut pas laisser de traces directement auprès du vendeur, mais il est prêt à le faire par l'intermédiaire d'une troisième partie, à la seule condition qu'il ait confiance que cette troisième partie ne va pas faire collusion avec le vendeur.

3. *Tolérance totale au tracking* : Le client n'a aucun problème à l'idée de laisser des

[3]En fait, cette technique utilise l'historique d'achats du client. Le vendeur a donc besoin de garder la trace de ses interactions avec le système, d'une connexion à l'autre.

traces, même par le vendeur. Ce niveau de protection est typiquement assuré à travers l'utilisation d'un même peusodnyme anonyme dont se sert le client auprès du vendeur d'une connexion à une autre.

4.2.2 Données du vendeur

Le vendeur participe au processus de vente à travers son catalogue de produits, sa boîte à outils décisionnelle (ex. : fonctions d'utilité[4], techniques de filtrage), etc. Le vendeur doit pouvoir protéger tout ou une partie de son catalogue, en fonction de son contenu et du type de client qu'il a en face de lui.

Idéalement, le vendeur voudrait garder son catalogue complètement secret pour se protéger contre la concurrence. Par exemple, il peut vouloir que les autres vendeurs ne prennent pas connaissance du contenu de son catalogue, surtout les prix de vente des produits. Toutefois, cela n'a aucun sens de garder le catalogue complètement secret. En effet, comment peut-on acheter auprès d'un vendeur qui refuse de dire si oui ou non il dispose du produit qu'on désire ? Dans ces conditions, nous avons pensé à séparer les données du vendeur en trois composantes, en fonction de leur sensibilité :

- *Données très sensibles* : ce sont des données que le vendeur juge très sensibles. Bien sûr, tout va dépendre du vendeur. Dans notre cas, nous considérons par exemple que le prix de vente et la stratégie utilisés dans la phase de négociation du prix de vente final d'un produit donné sont des données très sensibles (voir, **justification** dans la suite de cette section).
- *Données moyennement sensibles* : il s'agit des données que le vendeur décide de protéger mais qu'il accepte de dévoiler, à un rythme très lent toutefois, c'est-à-dire, à un rythme qui ne permettrait pas au client de tout apprendre en un laps de temps très court. En particulier, le vendeur ne souhaite communiquer qu'avec des humains et non des robots informatiques. Pour cela, il peut par exemple se servir des CAPTCHAS [180] (section 5.2) pour ralentir la cadence des requêtes du client. Les CAPTCHAS sont des programmes informatiques conçus pour générer des tests

[4]Utilisées dans la phase de négociation 1.7.

que les humains peuvent réussir, mais pas les machines. Les caractéristiques des produits (désignation, taille, couleur, marque, prix, garantie, etc.) constituent des exemples de données moyennent sensibles.

- *Données non sensibles* : ce sont des données qui ne nécessitent aucune protection de la part du vendeur. C'est le cas, par exemple, des données démographiques[5] du site Web commercial (adresse IP, adresse postale, etc.).

Justification : Il est intéressant de se pencher sur les raisons qui pousseraient un vendeur à considérer certaines données comme étant très sensibles. Pour cela, il suffit d'analyser l'importance de certains éléments qui sont à la disposition du vendeur :

- **Le prix** : il constitue la valeur résultant de toutes les autres caractéristiques d'un produit donné. Donc, à caractéristiques égales, seul le prix peut faire toute la différence entre deux vendeurs arbitraires. Nous pensons donc que le prix est le nerf de la concurrence et, par conséquent, il devrait être protégé.
- **La stratégie de négociation** : si le vendeur dévoile la stratégie qu'il utilise dans la phase de négociation du prix de vente final d'un produit, alors le client serait en mesure de toujours acheter au prix le plus bas, assorti des meilleures conditions de vente. C'est donc une autre donnée très sensible pour le vendeur.
- **Les comportements d'achats des clients** : les comportements d'achats sont surtout récupérés sous forme de feedback dans les systèmes de recommandation. Dans notre cas, sans nuire à la généralité, nous nous sommes limité à trois techniques, à savoir le filtrage par contenu, le filtrage collaboratif et le filtrage démographique. Dans les deux premiers cas, le client met généralement à la disposition du vendeur de l'information sur les produits qu'il a achetés dans le passé et, éventuellement, les votes qu'il a faits sur ces produits. Le vendeur peut ainsi faire le calcul de similarité et des prédictions en se basant sur cette information. Pour ce qui est du filtrage démographique, les clients fournissent leurs

[5]Ceci n'est pas toujours le cas ; certains vendeurs pourraient vouloir garder secrète une partie de leurs données démographiques en l'occurence leur emplacement physique, pour se protéger du terrorisme, par exemple.

données démographiques au vendeur. Ces données sont alors utilisées pour aider, d'une part, à la construction des catégories de clients et, d'autre part, à placer chaque client dans la catégorie qui correspond à ses données démographiques. Le vendeur doit donc être en mesure de protéger toutes ces données pour deux raisons principales : premièrement, ces données font partie de son *business* et, deuxièmement, le vendeur est responsable de la sécurité de toutes les données qu'il a colligées sur les clients qui lui ont fait confiance.

Nous allons maintenant présenter les mécanismes de protection de la vie privée que nous avons introduits grâce à la séparation des données décrite ci-dessus. Nous commençons par le nouveau modèle que nous avons proposé pour gérer le comportement et les transactions des clients en commerce électronique.

4.3 Création du modèle BCBB

Le modèle CBB n'apporte pas de réponses aux questions liées à la violation de la vie privée du client. À notre avis, il est important d'aider le client à faire des achats sans donner au vendeur de l'information qui pourrait par la suite mettre sa vie privée en danger. En particulier, il est question d'empêcher la formation de coalitions de vendeurs en vue d'une constitution de dossiers sur les clients. En partant du modèle CBB standard, nous avons éliminé toute possibilité chez le vendeur de colliger des données privées du client avant, pendant et après l'achat.

Cette démarche nous a conduit à la création d'un nouveau modèle : le BCBB (*Blind Customer Buying Behaviour*). Avant de donner un aperçu global de ce modèle, nous définissons d'abord quelques concepts utiles à sa compréhension.

4.3.1 Définitions

- ***Demande masquée*** : C'est une requête envoyée au vendeur par le client, avec la propriété que le vendeur est en mesure d'y répondre, sans pour autant être en mesure de prendre connaissance de son contenu de manière explicite. Les expressions *requête masquée* ou, dans la cadre de la négociation, *offre masquée* peuvent être utilisées en lieu et place de "demande masquée". La demande masquée s'oppose à la demande *en clair* en ce sens que la dernière donne au vendeur toute l'information qu'elle contient.

- ***Masquage*** : C'est une procédure qui permet de passer d'une demande, requête ou offre en clair à son *équivalent* masquée. Ceci se fait par voie de chiffrement (chapitre 2).

4.3.2 Notions préliminaires

La livraison anonyme des produits physiques et les systèmes de recommandation préservant la vie privée constituent des composantes essentielles à notre modèle BCBB. Nous allons donc les introduire dans cette section avant de passer, en section 4.3.3 à la synthèse globale du modèle.

4.3.2.1 Livraison anonyme des produits physiques

Une des limites de notre principe de masquage concerne la livraison de produits. Le cas des produits numériques a été analysé en section 3.9.4.1[6]. Pour ce qui est des produits physiques, comme par exemple la bague que veut acheter Sir Bob, nous avons fait intervenir plusieurs tiers, plus précisément, des agents de livraison, pour acheminer le paquet contenant le produit du vendeur jusqu'au client. Ces agents n'ont pas accès au contenu du paquet ; ils sont tenus de n'agir que sur son étiquetage et son emballage. L'idée d'avoir plusieurs tiers de confiance dans le processus de livraison se résume essentiellement en la mise sur pied d'une technique de mixage *à la Chaum* [54]. Au sortir de ce mixage, le client peut récupérer son paquet auprès du dernier agent de livraison avec une probabilité faible qu'il soit connu du vendeur. En effet, avec plusieurs tiers de confiance enroulés dans le processus de livraison, il faudrait une coalition d'un nombre considérable d'entre eux pour que le client soit connu du vendeur. Le détail de tout ceci

[6]En effet, la navigation anonyme sur Internet implique la livraion anonyme des produits numériques.

est donné en section 5.7 du chapitre 5, avec la notion de centre de livraison anonyme ou ADC (*Anonymous Delivery Centre*) [11, 12].

4.3.2.2 Systèmes de recommandation préservant la vie privée

Il faut parfois aller au-délà d'une simple recherche de produits dans le catalogue du vendeur à partir d'un certain nombre de caractéristiques choisies par le client. Pour cela, on peut par exemple se servir d'un système de recommandation de produits qui se base sur des techniques de filtrage (section 1.6.3). La recommandation de produits est très importante en commerce électronique compte tenu, par exemple, des besoins de personnalisation décrits en section 3.7. Il est donc important de proposer un système de recommandation qui préserve la vie privée des clients. Ceci n'est toutefois pas simple. Par exemple, un système de recommandation basé sur la technique de filtrage collaboratif a besoin des votes des clients pour être en mesure de faire des prédictions à l'intention du client courant. Comment faire si on ne veut pas exiger que plusieurs clients soient simultanément en ligne [49, 50] ou alors, si on ne veut pas avoir des prédictions *approximatives* car basées sur des perturbations aléatoires [133] ? Pour répondre à cette question, nous avons introduit la notion de *Tiers de semi-confiance* dans le processus de recommandation de produits au client. Un tiers de semi-confiance a les caractéristiques suivantes :

- Le client ne lui fait confiance que pour une partie de ses données privées. Par exemple, le tiers de semi-confiance pourrait disposer des données démographiques sur le client, juste le temps de le catégoriser, mais ne jamais savoir ce que le client achète. Le tiers de semi-confiance peut être logé dans la plateforme même du vendeur ou alors disposer d'une plateforme individuelle. Dans l'un ou l'autre cas, notre solution se base sur l'hypothèse qu'il ne fera pas coalition avec le vendeur.
- Le vendeur[7] ne veut pas du tout dévoiler le contenu de son catalogue. En effet, il doit bloquer la voie à toute forme de concurrence venant de l'extérieur comme de l'intérieur. Ainsi, même si le tiers de semi-confiance est en mesure de participer

[7] Il faut noter que nous mettons beaucoup plus l'accent sur la protection de la vie privée du client que sur celle des données sensibles du vendeur.

efficacement à la production de prédictions pour le client, il doit lui être impossible de prendre connaissance du catalogue du vendeur, pour quelque raison que ce soit.

Il faut noter que, de par sa nature, le commerce électronique se sert toujours d'un tiers de confiance d'une manière ou d'une autre. Une justification tient par exemple au fait qu'un site Web commercial se doit de disposer d'un certificat numérique (section 2.4.8.2) émis par une autorité de certification, donc un tiers de confiance !

4.3.3 Étapes du modèle BCBB

Contrairement au modèle CBB composé de six étapes (section 1.5), le modèle BCBB [8–10] consiste en quatre phases que nous présentons sommairement ci-dessous :

1. **Recherche masquée** : Cette phase regroupe les étapes 1 à 3 du modèle CBB. Elle permet, d'une part, au client d'identifier personnellement son besoin, de penser à des produits qui pourraient le satisfaire et de fouiller dans le catalogue du vendeur sans lui dire explicitement ce qu'il recherche. Pour cela, le client et le vendeur exécutent le protcole BliS (*Blind Search*) que vous avons développé (section 5.3). D'autre part, au cas où le client a une idée vague du produit qu'il lui faut, cette phase lui permet de se servir d'un système de recommandation (section 1.6) capable de lui proposer des produits pouvant satisfaire son besoin, tout en protégeant sa vie privée. Dans ce cas, nous avons mis sur pied le système de recommandation ALAMBIC [13, 14] (section 5.8). Le protocole BliS et le système ALAMBIC préservent tous les deux la vie privée du client ainsi que les données sensibles du vendeur.

2. **Négociation masquée ou BliN (*Blind Negotiation*) :** C'est l'équivalent de l'étape 4 du modèle CBB. Dans cette phase, le client est appelé à négocier (section 1.7) le prix de vente finale ainsi que les conditions de vente (ex. : la garantie) avec le vendeur sans que ce dernier sache de quel produit il s'agit, encore moins les prix de vente initial et final.

3. **Paiement masqué et livraison anonyme ou BliP (*Blind Payment and Delivery*) :** Cette phase fait référence à l'étape 5 du modèle CBB. Ici, le client est invité à

payer le montant correspondant au prix de vente final sans laisser de traces, c'est-à-dire sans que le vendeur sache qui a payé et combien. Similairement, il reçoit la livraison du produit acheté de manière anonyme, c'est-à-dire que le vendeur ne parvient pas à identifier le client à l'issue du processus de livraison.

4. **Service après vente masqué ou BliM (*Blind Maintenance*) :** Cette phase correspond à la phase 6 du modèle CBB. À travers elle, le vendeur est en mesure d'offrir le service après vente sans qu'il ait au préalable pris connaissance de l'identité ou des données démographiques du client.

Le modèle BCBB garantit la vie privée du client dans la mesure où il lui permet d'exécuter tout le processus d'achat avec le vendeur sans que ce dernier ait la possibilité de colliger des données personnelles sur lui. Les quatre phases du modèle BCBB sont décrites sous la forme de protocoles au chapitre 5.

4.4 Plateforme du client

Il serait erroné de penser à résoudre le problème de protection de la vie privée du client sans lui donner un moyen de lutte en local (sur sa machine) contre tous ces outils d'espionnage que nous avons présentés en section 3.6. En effet, que peut donc le masquage si, par exemple, le vendeur envoie au client, en même temps que la description du produit qu'il a sollicité, un Web bug pour l'épier sur sa propre machine ? Cette question témoigne de l'importance à créer un ensemble plus *complet* de mécanismes de protection de la vie privée. Un tel ensemble doit prendre en compte :

- Les *échanges visibles* : il s'agit des échanges normaux entre le client et le vendeur. Par exemple, Sir Bob peut demander clairement à Alice "Avez-vous des bagues à vendre ?" et recevoir la réponse "Oui, j'en ai".
- Les *échanges invisibles* : il s'agit essentiellement des données que le vendeur vient *chercher* sur la machine du client et cela, à son insu, en se servant par exemple de Spywares ou des Web bugs. Ceci inclut donc des installations par le vendeur de logiciels sur la machine du client, à son insu dans la plupart des cas. De tels logi-

ciels peuvent permettre par la suite au vendeur de communiquer avec la machine du client de manière illicite. Le logiciel de partage de musique Kazaa peut encore être cité en exemple ici, lui qui s'installe sur la machine du client en même temps que son spyware intégré *Cydoor*. Quoi que cette installation se fasse avec l'accord du client, il est toutefois improbable que celui-ci lise un tel accord avec de l'accepter (en cliquant sur un bouton "J'accepte", "Oui", etc.).

4.5 Conclusion

Nous avons consacré ce chapitre à la présentation sommaire de notre approche. Notre point de départ a été la description des données privées du client et des données sensibles du vendeur. En particulier, nous avons introduit la notion de SLP (Stop List for Privacy) ; une liste dans laquelle le client met ses données et applique des niveaux de tolérance (zéro, moyenne ou totale) à son contenu lors des transactions avec les vendeurs. Des niveaux de tolérance ont également été définis pour la composante "tracking" de la violation de la vie privée. De même, nous avons arrêté trois degrés de sensibilité pour les données du vendeur. Bien sûr, les nombres et types de niveaux de tolérance et/ou de degrés de sensibilité que nous avons choisis peuvent être revus : nous avons simplement voulu montrer la nécessité de classifier les données pour chaque entité et d'y appliquer une politique de gestion. En effet, à partir de la classification des données, nous avons mis en relief notre apport dans la lutte contre la violation de la vie privée en commerce électronique. Cet apport se résume en trois concepts principaux : le masquage, le tiers de semi-confiance et l'association de plusieurs tiers de confiance.

Le principe de masquage nous a permis de créer le modèle BCBB qui réduit le modèle CBB à quatre étapes, à savoir la *recherche masquée*, la *négociation masquée*, le *paiement masqué et la livraison anonyme* et la *service après-vente masqué*.

Toutefois, le masquage seul ne suffit pas. En particulier, il ne permet pas de faire des recommandations de produits tout en préservant la vie privée, encore moins ne permet-il de faire une livraison anonyme des produits physiques. Nous avons donc présenté les notions de tiers de semi-confiance et de centres de livraison anonyme pour permettre respectivement au client de recevoir des recommandations de produits et la livraison de produits physiques, sans craindre pour sa vie privée.

Nous avons enfin fait remarquer que toute solution au problème de la violation de la vie privée passe par la prise en compte de la machine du client, compte tenu des outils invasifs que nous avons décrits en section 3.6.

Ce chapitre était consacré à une présentation sommaire de notre approche. Nous sommes maintenant prêts à décrire, dans les détails, les différents concepts introduits ci-dessus. C'est ce qu'apporte le chapitre 5.

CHAPITRE 5

PROTOCOLES DU MODÈLE BCBB

Au chapitre 4, nous avons présenté, de façon sommaire, notre approche pour résoudre le problème de la violation de la vie privée en commerce électronique. Dans ce chapitre, nous donnons une description détaillée de cette approche.

5.1 Introduction

Il existe différents modèles et théories pour capturer le comportement d'achat d'un client en commerce électronique. On peut, par exemple, citer le modèle Nicosia [125], le modèle Howard-Sheth [93], le modèle Engel-Blackwell [78], le modèle Bettman [32], le modèle Andreasen [19] et, plus récemment, le modèle Guttman-Moukas-Maes [91] qui est considéré comme standard de nos jours, pour avoir réussi à unifier les différents modèles précédents.

Bien que différents, ces modèles partagent tous une liste semblable de six étapes fondamentales qui guident le comportement d'achat du client. Le modèle Guttman-Moukas-Maes est beaucoup plus adapté au commerce électronique avec l'utilisation des technologies basées sur les agents mobiles [17, 111].

Au chapitre premier, nous avons décrit le modèle Guttman-Moukas-Maes. Sans vouloir revenir sur cette description, un `constat` s'impose : *le modèle Guttman-Moukas-Maes ne tient pas compte de la vie privée du client.* À notre avis, la capture du comportement du client est le meilleur moyen dont peut se servir le vendeur pour fouiner dans sa sphère privée (section 4.2.1). C'est donc à ce niveau que nous avons pensé à intégrer les procédures de protection de la vie privée.

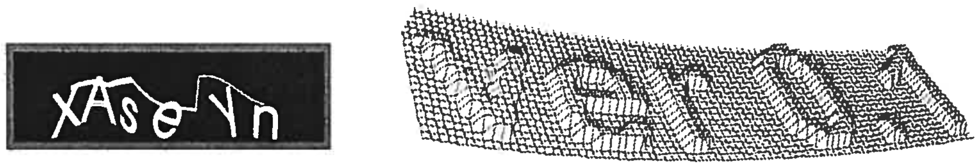
Nous avons consacré le chapitre 4 à la description sommaire des grandes lignes de notre approche, qui a donné naissance au modèle BCBB. Dans ce chapitre, les quatre phases du modèle BCBB sont décrites sous forme de protocoles que le client, le vendeur et éventuellement des tierces parties, sont invités à exécuter pour mener une transaction à terme. À titre de rappel, ces phases sont : (1) la recherche masquée (BliS), (2) la négociation masquée (BliN), (3) le paiement masqué et la livraison anonyme (BliP) et, (4) le service après vente masqué (BliM).

Voici l'organisation de ce chapitre. La section 5.2 regroupe quelques définitions et notations importantes dans la compréhension des protocoles du modèle BCBB. Les protocoles BliS, BliN, BliP et BliM sont par la suite décrits respectivement dans les sections 5.3, 5.4, 5.5 et 5.6. Viennent ensuite les centres de livraison anonyme (section 5.7) et les systèmes de recommendation préservant la vie privée (section 5.8). Enfin, nous faisons une synthèse du modèle BCBB (section 5.9), puis nous discutons de sa généricité, sa viabilité et son applicabilité en section 5.10, avant de clore le chapitre en section 5.11.

5.2 Préliminaires

En vue d'une meilleure compréhension des protocoles décrits dans ce chapitre, nous avons jugé utile de présenter ci-dessous certaines notations utilisées dans lesdits protocoles. Nous mentionnons déjà que nous nous sommes servi des cryptosystèmes à clé publique et/ou à clé secrète pour chiffrer les échanges entre le client et le vendeur, dépendamment du protocole. D'une manière générale, nous utilisons :

- Le *Problème du millionnaire* de Yao [184] : il s'agit d'une situation dans laquelle deux millionnaires veulent savoir lequel est le plus riche, sans que l'un apprenne le montant dont dispose l'autre. De nombreuses recherches sont menées pour trouver des solutions efficaces à ce problème [35, 38, 107, etc.]. En ce qui nous concerne, nous considérons que le client et le vendeur vont se servir d'une solution à ce problème pour comparer le prix maximal que le client peut payer et le prix minimal que le vendeur peut accepter. Ceci est fait au tout début de la négociation masquée (BliN).
- Le système de cryptographie ElGamal (section 2.4.2.2).
- Les CAPTCHAS : ce sont des programmes conçus pour générer des tests que les humains peuvent réussir, mais pas les machines. La figure 5.1 illustre des exemples



Valeur : **XAseYn** Valeur : **ver 0.1**

FIG. 5.1 – Exemples de CAPTCHAS.

de CAPTCHAS faits en 2D (à gauche) et 3D [1]. Nous signalons que les CAPTCHAS en 2D sont vulnérables tandis que ceux en 3D sont plus résistants (du moins, pour l'instant !). Le site Web [209] donne des détails sur les faiblesses des CAPTCHAS et propose ceux qui sont sensés être difficiles à briser. On y trouve d'ailleurs un état de l'art sur les CAPTCHAS, incluant leur utilisation, développement et *hacking*[2].

– Le *chiffrement de code* : c'est une technique qui vise à protéger le code mobile qui est exécuté dans des machines distantes possiblement suspectes (non dignes de confiance). De nombreux problèmes relatifs à la sécurité du code mobile en général, et à celle du code des agents mobiles en particulier, ont été présentés par Sander et Tschudin [152] en 1998. Par example, un agent mobile doit être capable de protéger l'intégrité et l'exécution de son code, utiliser ses clés sécrètes dans un environnement public pour faire des calculs, et assurer la protection du code (contre le plagiat par exemple) ainsi que les données internes (par exemple, toute ou une partie des données intermédiaires pendant l'exécution du code). Les auteurs de [152] ont aussi introduit la notion de calcul avec des fonctions chiffrées (*Computing with Encrypted Functions*). En particulier, ils ont proposé un solution générale au problème d'évaluation non-interactive des fonctions chiffrées (*non-interactive Evaluation of Encrypted Functions*) : "Alice dispose d'un algorithme pour calculer une fonction f. Sir Bob dispose d'une entrée x et souhaite calculer $f(x)$. Toutefois, Alice ne veut pas que Sir Bob dispose de l'information concer-

[1]Dimensions 2 et 3.
[2]Ici, il s'agit du bris des CAPTCHAS.

TAB. 5.1 – Obfuscation de code.

Code	Exemple d'obfuscation
int $i = 1$;	int $i = 11$;
while ($i < 1000$) {	while ($i < 8003$) {
... $A[i]$ ... ;	... $A[(i-3)/8]$ ... ;
i++ ;	i += 8 ;
}	}

nant f. De plus, Sir Bob ne devrait pas avoir à interagir avec Alice pendant le calcul de $f(x)$." [Nous avons changé Bob en Sir Bob dans cet énoncé que nous avons traduit en français]. Pour plus de détails, il est utile de lire [152].

– L'*obfuscation de code* : c'est un processus qui transforme un programme donné pour rendre son code très difficile à comprendre et donc plus résistant au *reverse engineering*[3]. Le code résultant de ce processus porte le nom de *code obfusqué* [64, 65]. Le tableau 5.1 est un exemple très simple illustrant un morceau code et sa version obfusquée.

  En pratique, le code obfusqué peut être le résultat d'une ou de plusieurs transformations de code. Malgré le fait qu'il diffère de manière significative du code original, le code obfusqué doit produire le même résultat que l'original, possiblement avec quelque lenteur dans l'exécution.

  Le chiffrement et l'obfuscation de code sont utilisés dans le protocole de négociation masquée (section 5.4) ainsi que dans le système ALAMBIC (section 5.8). Dans le premier cas, contrairement aux auteurs de [152], nous avons créé l'interaction entre le client et le vendeur pour permettre au premier d'évaluer la fonction d'utilité chiffrée et obsfuquée du second. Dans le second cas, le code de l'Agent Alambic (section 5.8) doit être chiffré et obfusqué, de sorte que le vendeur ne soit pas en mesure de prendre connaissance de son exécution (état interne, variables manipulées, etc.).

[3]Méthode consistant à reproduire le code source d'un programme à partir de son code exécutable.

5.3 Recherche masquée (BliS)

BliS [8] est un protocole d'achat/vente dans lequel les requêtes du client et les réponses du vendeur sont *aveuglées*, de sorte que :

- le vendeur n'apprend rien sur l'identité ou le besoin du client ;
- le client ne peut apprendre qu'un et un seul prix de vente de produit du catalogue du vendeur, pendant une session donnée. Une session est ouverte lors de la connexion du client sur le site commercial du vendeur et a une durée limitée.

Le fait pour le client de ne pouvoir apprendre qu'un seul prix de vente protège le catalogue du vendeur contre les concurrents qui pourraient se faire passer pour des clients et avoir un accès complet au catalogue. Par ailleurs, la protection des prix de vente fait appel à une considération plus ou moins philosophique. En effet, imaginons un monde dans lequel aucun vendeur ne connaît le prix de vente de ses concurrents. Comment les choses se passeraient-elles dans un tel monde ? Sans avoir la prétention de faire une analyse complète de cette question et en déduire une liste exhaustive de possibilités qui pourraitent se présenter, nous avons dégagé trois points :

- chaque vendeur vendrait au prix le plus bas possible (en fonction de son prix de revient) ;
- il n'y aurait pas de concurrence déloyale ;
- il n'y aurait pas de considérations sociologiques, du genre race, religion, gouvernement, etc.

D'une manière plus précise, à la fin de BliS, le client sait si le vendeur dispose du produit ayant fait l'objet de sa recherche ou d'un produit de remplacement. Il obtient alors une description complète dudit produit, incluant son prix de vente. Il peut aussi apprendre que le catalogue du vendeur ne contient pas le produit demandé, ou alors que ce produit fait partie d'une commande en cours.

Avant de décrire formellement BliS (section 5.3.3), nous allons d'abord présenter le mécanisme de recherche masquée, ainsi que la nécessité pour le vendeur de réorganiser sa base de données.

5.3.1 Mécanisme de recherche masquée

Dans une recherche masquée, le client n'envoie pas sa requête au vendeur d'un trait. Il divise plutôt sa requête en sous-requêtes qu'il envoie l'une après l'autre en tenant compte, dans l'envoi courant, des réponses qu'il a reçues précédemment. Ces réponses pourraient, par exemple, l'informer sur l'existence et la disponibilité d'un produit dans le catalogue. Dans le cas de Showbiz, si Sir Bob veut acheter une *bague en or de taille 2 munie d'une pierre en diamant*, il pourrait procéder comme suit : dans un premier temps, il masque la requête "Avez-vous une bague ?" et l'envoie à Alice. Le masquage, rappelons-le, signifie qu'Alice ne sera pas en mesure de prendre connaissance du contenu intelligible de la requête. Mais elle sera néanmoins en mesure de répondre "J'en ai en stock", si elle dispose effectivement de bagues à vendre. Par la suite, Sir Bob va continuer avec les requêtes masquées "Avez-vous une bague *en or* ?", "Avez-vous une bague en or *de taille 2* ?" et, finalement, "Avez-vous une bague en or de taille 2 *munie d'une pierre en diamant* ?". Requêtes auxquelles Alice repondrait par l'affirmative si elle dispose, bien sûr, d'au moins une bague répondant aux critères ainsi définis (toujours de manière masquée), par Sir Bob. Dès lors, si Sir Bob est satisfait par l'une des réponses qu'il a reçues d'Alice, ce n'est qu'à cet instant précis qu'il est autorisé à demander le prix de cette "bague en or de taille 2 munie d'une pierre en diamant". Une fois la demande de prix faite, il lui devient impossible de changer quelque caractéristique que ce soit. La contrainte de ne plus pouvoir changer de caractéristique permet de s'assurer que Sir Bob n'apprend qu'un seul prix de vente de produit dans le catalogue d'Alice pendant une session de recherche. En effet, tant que le prix de vente d'un produit n'est pas demandé, Alice laisse Sir Bob faire les ajustements qu'il souhaite sur les caractéristiques du produit devant satisfaire son besoin. Par exemple, si Alice ne dispose pas de bague munie de diamant, Sir Bob pourrait choisir une autre pierre (ex. : saphir).

En somme, une fois que le prix de vente a été demandé par Sir Bob, Alice lui répond et le mécanisme de recherche masquée pour la session en cours prend fin. Soit *prod* le produit choisi par Sir Bob. Ce dernier fait désormais face à un trois possibilités. Soit il accepte le prix de vente proposé par Alice. Soit il décide de négocier, avec Alice, le prix de vente final de $prod$, en se servant du protocole de négociation masquée décrit dans la section 5.4. Soit alors, il opte pour la fermeture de la session en cours et l'ouverture d'une autre session et ainsi engager un nouveau processus de recherche masquée.

5.3.2 Organisation du catalogue du vendeur

Nous supposons que le vendeur dispose d'une base de données $Y = [Y_1, Y_2, \cdots, Y_l]$ comprenant tous les produits à vendre (il y en l dans ce cas). Chaque produit Y_i, $1 \leq i \leq l$, est décrit par des caractériques, par exemple, le *nom*, la *largeur*, la *longueur*, la *hauteur*, la *couleur*, etc. Le client recherche donc un certain produit Y_i dans le catalogue Y. Sans nuire à la généralité, nous supposons que chacun des produits est décrit par un ensemble *fixe*, $C = \{C_1, \cdots, C_m\}$, de m caractéristiques publiquement connues. En d'autres termes, ces m caractéristiques constituent des colonnes qui décrivent chaque produit Y_i de Y. Nous avons par ailleurs considéré deux caractéristiques spéciales : l'*état* du produit et l'*étiquette* de prix. L'état du produit est un message explicite qui donne l'information sur la disponiblité du produit correspondant à chaque ligne de la base de donnée Y. Un tel message pourrait être : *J'en ai en stock, Je ne vends plus ce produit, Ce produit a été remplacé par [...], Ce produit me sera livré dans une semaine*, etc. L'étiquette de prix contient, quant à elle, le prix de vente initial, le prix de vente minimum chiffré et un lien pour télécharger la stratégie de négociation sous forme de programme exécutable. (Ces deux dernières données vont exclusivement être utiles dans la phase de négociation de la section 5.4).

En clair, le catalogue du vendeur est formé des l lignes de la base de données Y et de $m+2$ colonnes issues des m caractéristiques de C et des caractéristiques spéciales "état" et "étiquette". Quand Sir Bob veut vérifier la disponibilité et le prix d'un produit donné dans le catalogue d'Alice, il choisit au préalable les caractéristiques qui l'intéressent et en forme un ensemble $\Delta \subseteq \{1, 2, \ldots m\}$, c'est-à-dire qu'il met $i \in \{1, 2, \ldots m\}$ dans Δ s'il est intéressé par la caractéristique C_i. Il dit explicitement à Alice la constitution de l'ensemble Δ. Alice forme alors une vue plus restreinte de son catalogue, formée uniquement des caractéristiques issues de Δ ainsi que des colonnes "état" et "étiquette". Compte tenu de la réduction du nombre de caractéristiques, il se peut que la colonne "état" affiche des contenus différents pour deux lignes pourtant identiques si on réduit la table aux colonnes (caractéristiques) issues de Δ. Par conséquent, Alice doit changer les messages contenus dans la colonne "état" suivant le principe suivant, appélé *Procédure de Remplacement d'État* (PRÉ) :

> **Principe PRÉ** : Si les contenus de deux lignes restreintes aux colonnes issues de Δ sont identiques, mais que la colonne "état" a des messages[4] différents, il faut alors choisir le message le plus à même de satisfaire le client. Par exemple, si la colonne "état" contient les messages "Produit en commande" et "Je l'ai en stock", alors Alice ne devra retourner qu'une ligne (au lieu de deux) contenant le dernier message (Je l'ai en stock).

Par ailleurs, si deux lignes diffèrent uniquement dans la colonne "prix", alors Alice choisit la ligne qui lui rapporte un bénéfice plus important, puisqu'il y va de ses intérêts de vendeur. Bien sûr, ces intérêts peuvent plutôt viser par exemple la vente de certaines catégories de produits, des promotions à faire, etc. Dans tous les cas, ce sera au vendeur de décider de la ligne à présenter au client, modulo sa politique et les circonstances de vente.

Dans le cas Showbiz, l'ensemble Δ pourrait contenir les caractéristiques *nom*, *matière*, *taille*, etc. Le mécanisme de recherche masquée permet alors à Sir Bob de savoir s'il existe une ligne dans le catalogue d'Alice qui sied à cet ensemble Δ et, si tel est le cas, il obtient l'"état" correspondant. Si une telle ligne n'existe pas ou s'il n'est pas satisfait, alors il peut soit raffiner son ensemble Δ en y ajoutant de nouvelles caractéristiques, soit mettre à jour les valeurs affectées à ses caractérisques (ex. : changement de la coupe de la pierre, d'*ovale* à *triangle*), soit changer complètement le contenu de l'ensemble Δ, soit alors décider de quitter Alice sans faire d'achat. Même après avoir changé d'ensemble, il se peut qu'à un moment donné, Sir Bob ait vu quelque chose d'intéressant et qu'il ait toutefois voulu passer en revue d'autres produits en espérant trouver mieux, mais qu'en fin de compte, il décide de faire marche arrière pour sélectionner un produit

[4] Ces messages sont préfabriqués par le vendeur.

considéré plus tôt.

Dans tous les cas, c'est la demande de prix par Sir Bob qu'Alice gère avec plus de rigueur, compte tenu du fait que cette donnée lui est *très sensible* (section 4.2.2). En effet, quel que soit ce que Sir Bob fait pendant une session, Alice ne lui permettra l'option de demande de prix qu'une et une seule fois. Cependant, des CAPTCHAS seront employés pour empêcher des attaques de robots sur les données moyennement sensibles d'Alice. Par exemple, nous avons opté pour un CAPTCHA chaque fois que Sir Bob change son ensemble Δ et à des intervalles réguliers quand il met cet ensemble à jour. Toutefois, il n'y a aucun besoin de CAPTCHA quand Sir Bob raffine sa question, puisque ce serait finalement trop ennuyeux pour un humain.

5.3.3 Formalisation de BliS

Soit $\mathcal{D}$, un *ensemble universel* permettant de décrire n'importe quel produit. En d'autres termes, $\mathcal{D}$ est un ensemble constitué de toutes les caractéristiques dont on peut se servir pour décrire un produit quelconque. $\mathcal{D}$ a donc pour éléments : *nom, couleur, matériel, largeur, longueur, hauteur, poids,* etc. Chaque caractéristique peut prendre une ou plusieurs valeurs ; par exemple, la caractéristique *couleur* peut avoir, entre autres, les valeurs *vert*, *rouge*, *jaune*. Dans le même sens, une caractéristique peut être simple ou composée. Dans le cas d'une caractéristique simple, un seul attribut est considéré. La longueur est un exemple de caractéristique simple. Une caractéristique composée quant à elle suppose l'existence de plus de deux attributs ; par exemple, le composition d'une bague pourrait être : 60% *or* et 40% *argent*.

Nous avons introduit la notion de *Codification Standard Universelle* dans le but d'uniformiser les valeurs que peut prendre une caractéristique. Le client se sert de cette codification pour décrire le produit dont il a besoin. Il applique ensuite une fonction de hachage H, publique (connue du vendeur et du client), sur les codes des valeurs des caractéristiques contenues dans Δ. Plus précisément, elle prend en entrée une concaténation des valeurs des caractéristiques et produit en sortie un haché compris entre 1 et N, où N est un entier suffisamment grand afin de réduire le nombre de collisions. Il est important de noter que nous n'avons pas besoin ici d'une fonction de hachage cryptographique car la sécurité de BliS ne dépend pas de la difficulté d'inverser H ou de trouver des préimages distinctes pour un certain point de l'ensemble d'arrivée. Dans la suite de ce chapitre, H est appelée *fonction d'indexation universelle*. La sortie de cette fonction est appelée *index* de recherche relatif à Δ et est notée $H(\Delta)$. De son côté, le vendeur crée une vue, $V(\Delta)$, relative à l'ensemble Δ. Les colonnes de cette vue sont exactement les caractéristiques présentes dans Δ. Ainsi, pour chaque ligne, ℓ, de $V(\Delta)$, le vendeur applique H sur les codes des valeurs des différentes colonnes, et obtient un digest $H(V(\Delta).\ell)$.

Dès lors, la requête du client se réduit à : "Donnez-moi l'information sur la disponibilité du produit d'index $H(\Delta)$". Toutefois, cette requête est transmise au vendeur de manière masquée tel que décrit dans la section 5.3.4. Signalons que le vendeur dispose des index permanents permettant d'identifier chaque ligne de son catalogue. Toutefois, la vue $V(\Delta)$ qu'il a créée et qui dépend de Δ contient de nouveaux index temporaires dont lui seul maîtrise la correspondance avec les index permanents.

Dans toute la section sur BliS, les variables t et m, ainsi que l'ensemble $\Delta = \{\delta_1, \delta_2, \cdots, \delta_{|\Delta|}\}$, $|\Delta| \leq m$, $\delta_i \in \{1, 2, \cdots, m\}$, sont telles que définies ci-dessus. Après que le vendeur ait reçu l'ensemble Δ et choisi les t_Δ lignes de son catalogue qui forment $V(\Delta)$, nous notons par $d_{\ell,j}$ la valeur contenue dans la ligne ℓ et la colonne δ_j, avec $1 \leq \ell \leq t_\Delta$ et $1 \leq j \leq |\Delta|$, de la table $V(\Delta)$. Pour chaque ligne ℓ de $V(\Delta)$, l'index obtenu par le vendeur est $H(d_{\ell,1}, \cdots, d_{\ell,|\Delta|})$. Si c_i désigne le choix du client pour la caractéristique C_{δ_i} (ex. : C_{δ_i} pourrait être "couleur" et c_i pourrait être le code universel pour le "rouge"), alors l'index de recherche du client est $H(c_1, \cdots, c_{|\Delta|})$.

5.3.4 Étapes du protocole BliS

Nous décrivons ci-dessous, en détail, le déroulement du protocole BliS, étape par étape. Nous rappelons que BliS est un protocole dont se sert Sir Bob pour mener une recherche masquée dans le catalogue d'Alice.

Initialisation : Nous rappelons que la fonction de hachage H introduite dans la section 5.3.3 donne en sortie des entiers compris entre 1 et N. Une fois pour toute, le vendeur choisit un nombre premier $q \geq N$ tel que $p = 2q + 1$ est aussi premier. À partir de cet instant (jusqu'à la fin de cette section), tous les calculs sont effectués dans $\mathbb{Z}_p^*$ et nous identifions les entiers entre 1 et $p-1$ comme éléments de $\mathbb{Z}_p^*$ quand cela est approprié. Soient $\mathcal{G}$ le groupe multiplicatif des résidus quadratiques dans $\mathbb{Z}_p^*$ et $\mathcal{R}$ l'ensemble des entiers compris entre 1 et q. Le vendeur choisit un générateur g de $\mathcal{G}$ et un entier aléatoire $x \in \mathcal{R}$, puis il calcule $s = g^x$. Pour clore l'étape d'initialisation, le vendeur rend publiques les valeurs p, g et s (et implicitement, $q = (p-1)/2$), pour qu'elles soient utilisées par n'importe quel client. Plus précisémet, le vendeur dépose sa clé publique $[p, g, s]$ auprès d'une Autorité de Certification ou CA (Section 2.4.8.2), qui lui fournit alors un cerificat d'authentification. Toutefois, le vendeur garde x secret.

Dans ce qui suit, nous supposons que le "Decision Diffie-Hellman (DDH) Assumption" est valide dans $\mathcal{G}$ (Section 2.4.2.1) et, par conséquent, que le cryptosystème ElGamal basé sur $\mathcal{G}$ est sémantiquement sécuritaire (Section 2.4.2.2). Supposons pour simplifier que les prix peuvent être représentés par des entiers compris entre 0 et $\mathcal{P}-1$ pour un certain entier $\mathcal{P}$, qui est le même pour tous les vendeurs. En plus de la condition que $q \geq N$, nous supposons aussi que q a été choisi suffisamment grand de sorte qu'Alice peut encoder n'importe quelle paire $\langle m, w \rangle$ comme élément de $\mathcal{R}$, où m encode l'*état* du produit, c'est-à-dire sa disponibilité (Section 5.3.2) et w désigne son prix.

Chaque fois que le client décide de faire des achats, il demande à voir le certificat à clé publique $[p, g, s]$ du vendeur et vérifie qu'il est légitime. (Dans certains cas, ceci peut nécessiter de l'interaction avec le CA). Ensuite, le client choisit une clé symétrique aléatoire pour des besoins d'authentification et l'envoie au vendeur *via* le cryptosystème ElGamal (en utilisant la clé publique du vendeur). Le but de cette clé d'authentification sera d'assurer l'intégrité de tous les messages futurs entre le vendeur et le client. Ceci sert à prévenir une attaque *Person-in-the-middle*, issue par exemple d'un compétiteur sans scrupules qui voudrait miner la crédibilité du vendeur aux yeux du client en se plaçant furtivement entre les deux (vendeur et client) après la poignée de main initiale. Veuillez prendre en note que seule l'authentification est désirable ici, pas la confidentialité, car le vendeur ne dit que des choses que chacun peut obtenir en interagissant de manière honnête avec lui, et les messages du client n'ont de sens que pour lui-même. Bien sûr, on aura besoin de la confidentialité plus tard, quand viendra le temps de payer en utilisant le protocol BliP (Section 5.5), mais ceci ne nous intéresse pas pour le moment.

Après la phase d'initialisation que nous venons de décrire, les étapes suivantes sont exécutées entre le vendeur et le client :

1. Le vendeur choisit aléatoirement un entier $\wp$ compris entre 0 et $\mathcal{P} - 1$, qu'il garde secret pour l'instant. Cet entier servira à déchiffrer les prix. Comme cette étape est reprise plusieurs fois, une liste constituée des différents $\wp$ choisis, à chaque passage par cette étape, est gardée secrète pour une utilisation ultérieure.

2. Le client choisit un ensemble Δ des caractéristiques à partir desquelles il souhaite formuler sa requête. Il annonce l'ensemble Δ au vendeur. Pour chaque colonne $\delta_i \in \Delta$, il choisit en secret la codification v_i qui correspond à la valeur recherchée pour la caratéristique C_{δ_i}. Il applique la fonction d'indexation universelle H et obtient l'index de recherche $u = H(v_1, \cdots, v_{|\Delta|})$. Il choisit aléatoirement $b \in \mathcal{R}$, formule sa requête masquée $y = u^2 \cdot g^b$ et l'envoie au vendeur.

3. En se servant de la PRE (Section 5.3.2), le vendeur crée la vue $V(\Delta)$ correspondant à l'ensemble Δ annoncé par le client. Cette vue contient $t_\Delta \leq t$ lignes à travers lesquelles le client recherche un produit. Le vendeur applique H sur chacune des lignes ℓ de $V(\Delta)$ et obtient un index $e_\ell = H(d_{\ell,1}, \cdots, d_{\ell,|\Delta|})$, $\ell = 1, \cdots, t_\Delta$. Pour chaque ligne ℓ, désignons par m_ℓ et w_ℓ le message que contient la colonne "état" et son étiquette de prix, respectivement. Le vendeur chiffre les prix en $z_\ell = \wp + w_\ell \bmod \mathcal{P}$ et envoie les messages $M_\ell = \langle m_\ell, z_\ell \rangle$ au client, $l = 1, \ldots, t_\Delta$, mais sous forme masquée :

$$E_\ell = \left((y/e_\ell^2)^x \,,\; g^{a_\ell} \,,\; M_\ell^2 \cdot (y/e_\ell^2)^{a_\ell}\right),$$

   où a_ℓ est choisi aléatoirement dans $\mathcal{R}$, de manière indépendante pour chaque $\ell = 1, \ldots, t_\Delta$. Le vendeur envoie E_ℓ au client, $\ell = 1, \ldots, t_\Delta$.

4. Le client a reçu du vendeur t_Q triplets $(\alpha_\ell, \beta_\ell, \gamma_\ell)$, mais tout au plus un seul parmi eux est significatif pour sa requête : il s'agit de celui qui correspond à ℓ tel que $u = e_\ell$, si un tel ℓ existe. Pour le retrouver, il suffit de remarquer que $\alpha_\ell = (y/e_\ell^2)^x = (u^2 \cdot g^b/e_\ell^2)^x$, et que ceci conduit à $g^{bx} = s^b$ si et seulement si $u^2 = e_\ell^2$. Ainsi,

le client calcule s^b une fois pour toute, puis il fait défiler tous les triplets reçus jusqu'à ce qu'il tombe sur celui qui est tel que $\alpha_\ell = s^b$. À partir de cet instant, le client calcule :

$$\gamma_\ell \cdot (\beta_\ell^b)^{-1} = M_\ell^2 .$$

Il obtient $M_\ell = \langle m_\ell, z_\ell \rangle$ en calculant la racine carrée de M_ℓ^2 qui est un élément de $\mathcal{R}$ (Section 2.4.2.2). À présent, il examine le message m_ℓ, qui lui donne l'information sur la disponibilité du produit qu'il recherche. Dès lors, le client fait face à un triple choix :

(a) Il peut souhaiter continuer sa recherche dans le catalogue du vendeur, soit parce qu'il n'est pas satisfait par l'offre courante ou alors parce qu'il pense pouvoir trouver quelque chose de mieux. Dans ce cas, il réajuste l'ensemble Δ et le protocole est relancé à partir de l'étape 1, possiblement, après l'avoir soumis à un CAPTCHA [180].

(b) S'il est satisfait du résultat de sa recherche, soit parce qu'il est intéressé par ce qu'il vient de trouver à la ronde en cours ou à l'une des précédentes rondes, il se sert du transfer inconscient pour obtenir du vendeur la valeur du $\wp$ qui a été utilisée pendant la ronde qui a finalement retenu son attention. Cela lui permet de déchiffrer le prix correspondant : $w_\ell = z_\ell - \wp \bmod \mathcal{P}$. À ce point-ci, le client peut soit dire au vendeur qu'il n'est plus intéressé à acheter, soit décider d'acheter (exécution du protocole de paiement masqué et de livraison anonyme—BliP), soit opter pour une négociation masquée—BliN—du prix de vente final.

(c) Il peut enfin considérer qu'il n'est pas au bon endroit, auquel cas il informe le vendeur de son départ et les deux mettent fin à l'exécution du protocole.

5.4 Négotiation masquée

Dans le modèle CBB, quand arrive le temps de négocier, le client et le vendeur se servent de *stratégies de négociation* (section 1.7) pour faire des choix et prendre des décisions. Dans cette section, nous parlons de la négociation entre le vendeur et le client sur les termes de la transaction électronique qui les lie. Les termes considérés ici sont le prix de vente du produit et possiblement quelques autres aspects tels que les conditions de la vente, la garantie, etc.

Comment peut-on alors négocier dans un environnement qui vise la préservation de la vie privée du client ? Dans la phase de recherche du produit (BliS), le vendeur n'a pas été autorisé à prendre connaissance du produit sélectionné par le client. Nous voulons que, dans cette phase de négociation, cette ignorance soit maintenue. En d'autres termes, nous demandons au client de négocier avec le vendeur le prix de vente d'un produit, mais ceci doit se faire sans que le vendeur sache son propre prix de départ, encore moins le produit qui fait l'objet de la négociation ![5] Par ailleurs, le vendeur ne doit jamais apprendre quoi que ce soit au sujet des offres faites par le client ou le prix de vente final qui sera éventuellement arrêté. Nous avons qualifié ce type de négociation de masquée compte tenu des propriétés que nous venons d'évoquer, et avons introduit le protocole de négociation masquée BliN (Blind Negotiation) [9].

Nous allons à présent donner le détail de BliN et illustrer son fonctionnement à travers l'exemple `Showbiz`. Supposons que c'est la qualité du diamant qui se trouvera sur la bague qui importe le plus aux yeux de Sir Bob. Ce dernier va donc mettre l'emphase sur les caractéristiques du diamant. Il faut savoir qu'il y a quatre aspects [196], connus sous le nom de *quatre* C, qui caractérisent un diamant : *Coupe–Carat–Clarté–Couleur*. Ces quatre C sont les attributs les plus importants lors du choix d'un diamant. Les valeurs possibles pour la Coupe sont : *brillant, ovale, radiant, marquise, cœur, émeraude, poire, princesse, triangle*, etc. La caractéristique Carat fait allusion au poids du diamant. La Clarté donne l'information sur les imperfections intérieures et extérieures du diamant ; ses valeurs possibles sont (de zéro imperfection à plusieurs failles) : FL, IF, VVS1, VVS2, VS1, VS2, SI1, SI2, SI3, I1, I2, I3. Quant à la Couleur, ses valeurs vont de D (la meilleure catégorie) à Z (la plus mauvaise catégorie). Pour illustrer notre paradigme de négociation masquée, nous allons considérer les attributs suivants : Prix, Coupe, Carat,

[5] Puisque le choix du produit à acheter est issu d'une recherche masquée.

Clarté, Couleur et Garantie.

5.4.1 Mécanisme de négociation masquée

Dans cette section, nous présentons les différentes composantes du protocole BliN. Désignons encore par P le produit choisi par le client à la fin du protocole BliS et qui fait à présent l'objet de la négociation. Soit $\mathcal{A}$ l'ensemble formé d'attributs de P, comme par exemple le "Prix" (à ne pas confondre avec le prix du produit, qui est en fait la *valeur* de l'attribut "Prix"). Ces attributs (pas leurs valeurs) sont connus du client et du vendeur. Si t désigne la cardinalité de $\mathcal{A}$, alors nous pouvons écrire $\mathcal{A} = \{A_1, \cdots, A_t\}$. Ainsi, dans l'exemple `Showbiz`, nous avons : $\mathcal{A} = \{Prix. Coupe, Carat. Clart. Couleur, Garantie\}$. Par ailleurs, soit $\mathcal{O}$, l'ensemble des tuples des valeurs que le vendeur et le client pourraient affecter aux attributs en cours de négociation. En d'autres termes, $\mathcal{O}$ est l'ensemble offres et contre-offres. Une offre (ou une contre-offre) $O \in \mathcal{O}$ est écrite sous la forme $O = (index; v_1; \cdots; v_t)$, où *index* désigne l'index du produit, P, choisi par le client à la fin de BliS (voir Section 5.3.4), et v_i est la valeur choisie pour l'attribut A_i, $i = 1, \cdots. t$. Dans le cas `Showbiz`, nous pouvons, par exemple, écrire O_s =(101 ; 1,5M ; brillant ; 1,12 ; SI ; H ; 5 ans).

La *Fonction d'Utilité* est un outil qui sert à évaluer l'utilité d'une offre donnée. Il existe plusieurs fonctions d'utilité. Dans la cadre de notre travail, nous ne nous intéressons pas à la formalisation mathématique de telles fonctions (au besoin, le lecteur pourrait consulter [98] pour plus de détails). Soit U_v la fonction d'utilité du vendeur. Dans un environnement sans négociation masquée, le vendeur devrait connaître sa propre fonction U_v, et le client lui enverrait en clair son offre ou sa contre-offre $O_c = (index; v_1; \cdots; v_t)$. Le vendeur évaluerait alors $u_c = U_v(O_c)$ et utiliserait cette valeur pour prendre une décision. En revenant sur l'exemple `Showbiz`, Alice évaluerait $u_c = U_v(O_s) = U_v$(101 ; 1,5M ; brillant ; 1,12 ; SI ; H ; 5 ans).

Dans un contexte de négociation masquée, le problème est plus complexe : le vendeur n'a d'information ni sur sa propre fonction d'utilité, ni sur l'offre faite par le client. Néanmoins, pour que la négociation ait lieu, le vendeur crée et assure la maintenance d'une table M, composée de $t + 3$ colonnes. Les colonnes allant de 1 à t correspondent

TAB. 5.2 – Showbiz : Table de négociation d'Alice.

Prix	Coupe	Carat	Clarté	Couleur	Garantie	Index	UVal	Promo
\$1,90M	Ovale	2,00	VVS1	E	3 ans	103	5	
\$1,75M	Coeur	1,86	FL	D	5 ans	101	4	P3F2
\$2,50M	Princesse	1,66	IF	E	À vie	104	2	
\$1,02M	Perle	0,96	SI1	H	3 ans	102	3	P1HALF2
\$1,83M	Coeur	1,46	IF	D	3 ans	101	4	P1FREE
\$0,65M	Coeur	0,46	VVS1	H	3 ans	101	3	

Remarques : Coupe, Carat, Clarté et Couleur constituent les quatre C du diamant, Index identifie le produit. UVal est la valeur d'utilité de l'offre et Promo renvoie aux offres promotionnelles.

aux attributs de la négociation. Les colonnes $t+1$, $t+2$ et $t+3$ sont respectivement l'index de P, la valeur d'utilité (UVal dans le tableau 5.2) comprise, dans notre cas, entre 1 et 5, et les promotions éventuelles (*Promo* dans le tableau 5.2) faites sur P (ex. : *P3F2 : Trois pour le prix de deux*, ou *P1HALF2 : À l'achat d'une unité, obtenez la deuxième à moitié prix*). Nous supposons que le vendeur ait une stratégie (ou un appareil) qui lui permette de produire des offres pour tous les produits de son catalogue. Ainsi il importe peu qu'il maintienne ses offres dans une table ou qu'il laisse plutôt l'appareil les produire sur demande. Ce qui compte est la valeur d'utilité que le vendeur affecte à une offre, quelle qu'elle soit. Par conséquent, nous considérons que la fonction d'utilité du vendeur reçoit en entrée l'index du produit (colonne $t + 1$) et l'offre (colonnes 1 à t), puis elle produit la valeur d'utilité correspondante. Le tableau 5.2 présente une vue partielle de M dans le cas Showbiz.

En se rappelant que notre objectif principal est de protéger la vie privée du client, il est clair que s'il envoie son offre en clair au vendeur, ce dernier prend connaissance du produit qui fait l'objet de négociation. Et dès lors, la vie privée du client peut se trouver menacée. Par conséquent, c'est au client d'évaluer la fonction d'utilité du vendeur, U_v^f. Mais il doit le faire de manière masquée (c'est-à-dire qu'il doit être difficile pour le client d'obtenir plus d'information sur la fonction U_v). Pour cela, le vendeur cache la fonction U_v à l'intérieur d'un programme, en se servant du schéma de *Chiffrement et d'Obfuscation de la Fonction d'Utilité* ci-dessous :

1. le vendeur chiffre sa fonction d'utilité U_v et obtient $E^v_{K_v} \circ U_v$, où E^v désigne l'algorithme de chiffrement E^v du vendeur et K_v sa clé secète ;

TAB. 5.3 – Requête : "Index=101 et UVal=4".

Prix	Coupe	Carat	Clarté	Couleur	Garantie	Index	UVal	Promo
1.75M$	Coeur	1.86	FL	D	5 ans	101	4	P3F2
1.83M$	Coeur	1.46	IF	D	3 ans	101	4	P1FREE

2. le vendeur crée un programme $\mathcal{P}(E^v_{K_v} \circ U_v)$;
3. le vendeur fait l'obfuscation du code de $\mathcal{P}(E^v_{K_v} \circ U_v)$ et obtient $\mathcal{P}_o = \mathcal{P}_o(\mathcal{P}(E^v_{K_v} \circ U_v))$.

C'est à la fin de BliS, lorsque le client demande le prix de vente du produit qu'il a choisi en secret, qu'il reçoit également la fonction d'utilité du vendeur, chiffrée et obsfusquée, $\mathcal{P}_o$.

Dans l'écriture ci-dessous, K_v, K'_v et K''_v désignent les clés secrètes du vendeur, K_c celle du client, alors que E^v et E^c sont respectivement les algorithmes de chiffrement du vendeur et du client. Il est nécessaire que ces deux algorithmes commutent, c'est-à-dire que $E^v_{K_v}(E^c_{K_c}(x)) = E^c_{K_c}(E^v_{K_v}(x))$ pour une quelconque donnée x. Pour faire une offre O_c, le client exécute $\mathcal{P}_o$ sur l'entrée O_c et obtient une valeur d'utilité chiffrée $\mathcal{P}_o(O_c)$. Sans perte de généralité, nous pouvons écrire $\mathcal{P}_o(O_c) = E^v_{K_v}(u'_v)$, puisque chiffrer dépend seulement de $E^v_{K_v}$ (il faut se rappeler ici que l'obfuscation d'un programme ne modifie pas sa sortie). Ici, u'_v est une valeur intermédiaire que le vendeur va transformer en u_v après avoir interagi avec le client comme suit : le client chiffre $E_{K_v}(u'_v)$ et envoie le résultat $E^c_{K_c}(E^v_{K_v}(u'_v))$ au vendeur. Puis, le vendeur déchiffre ce résultat et le chiffre de nouveau en $E^v_{K'_v}(E^c_{K_c}(u'_v))$. Il renvoie cette valeur au client. Finalement, le client obtient $u_v = E^v_{K'_v}(u'_v)$. Il est important de noter que l'interaction est nécessaire entre le client et le vendeur pour que le premier obtienne u_v. Si le client pouvait calculer u_v tout seul, alors il pourrait imaginer toutes sortes d'offres et le vendeur ne tirerait aucun bénéfice de la négociation. Il est également important que le vendeur n'apprenne pas u_v car, sinon, il pourrait fabriquer une fausse fonction d'utilité qui lui permettrait d'obtenir l'information sur le produit choisi par le client.

Pour illustrer ce processus, revenons à l'exemple `Showbiz`. Supposons que BliS a permis à Sir Bob de choisir une bague et d'en obtenir l'index, 101, le prix de vente, 2M\$, le meilleur prix sous une forme chiffrée, $E^v_{K''_v}(1.80\text{M\$})$, et la fonction d'utilité obfusquée et chiffrée, $\mathcal{P}_o$. Si le prix maximum que peut payer Sir Bob est $1,85$M\$, alors il lui faut négocier un meilleur prix avec Alice. Mais, il doit au préalable examiner (avec Alice) si cela vaut la peine de négocier. En d'autres termes, il doit s'assurer que son prix maximum est plus grand ou égal au meilleur prix que peut accepter Alice (pour ce produit d'index 101). Si cette assurance a lieu, comme c'est la cas dans notre exemple ($1.85\text{M\$} \geq 1.80\text{M\$}$), alors Alice et Sir Bob exécutent le protocole BliN. Supposons par exemple que la première offre de Sir Bob est O_c = (101 ; 1,7M ; Coeur ; 2,46 ; FL ; D ; 5 ans). Sir Bob évalue son offre en utilisant $\mathcal{P}_o$ et, avec l'aide d'Alice, obtient $u_v = 4$ (par exemple). Il interroge alors, toujours d'une façon masquée, le tableau 5.2 pour en obtenir les lignes correspondant à la requête "Index=101 et à UVal=4". Le résultat de cette requête est présentée dans le tableau 5.3. Sir Bob peut alors choisir une ligne du tableau 5.3, faire une contre-offre (probablement en tenant compte de tableau 5.3) ou mettre fin au processus de négociation.

5.4.2 Étapes du protocole BliN

Rappel : Quand, à la fin de BliS, le client fait la demande de prix du produit qu'il a sélectionné auprès du vendeur, il obtient :

- en clair, le prix de vente proposé par le vendeur : P_v ;
- chiffré, le programme qui implémente la fonction d'utilité du vendeur : $\mathcal{P}_o$;
- chiffré, le prix minimum que peut accepter le vendeur : $E^v_{K''_v}(P^v_{min})$, où E^v et K''_v sont respectivement l'algorithme de chiffrement et une clé privée du vendeur.

À cet instant (fin de BliS), le client obtient aussi l'information complète sur les attributs du produit qu'il a choisi. Ceci inclut l'index de ce produit dans le catalogue du vendeur.

De manière formelle, le protocole BliN fonctionne comme suit :

Initialisation[6] : Nous supposons que les index des produits et les valeurs d'utilité (Section 5.4.1) sont compris entre 1 et N, pour N suffisamment grand pour pouvoir couvrir

[6]Ceci est presque identique à la phase d'initialisation de BliS. Toutefois, nous avons tenu à simplifier la tâche au lecteur en ne remettant ici que ce qui est essentiel à la compréhension de BliN

tous les produits. Une fois pour toute, le vendeur choisit un nombre premier $q \geq N$ tel que $p = 2q + 1$ est aussi premier. Tous les calculs ci-dessous sont effectués dans $\mathbb{Z}_p^*$ et nous identifions les entiers entre 1 et $p - 1$ comme éléments de $\mathbb{Z}_p^*$ quand cela est approprié. Soient $\mathcal{G}$ le groupe multiplicatif des résidus quadratiques dans $\mathbb{Z}_p^*$ et $\mathcal{R}$ l'ensemble des entiers compris entre 1 et q. Le vendeur choisit un générateur g de $\mathcal{G}$ et un entier aléatoire $x \in \mathcal{R}$, puis il calcule $s = g^x$. Le vendeur rend publiques les valeurs p, g et s (et implicitement, $q = (p - 1)/2$), pour qu'elles soient utilisées par n'importe quel client. Plus précisémet, le vendeur dépose sa clé publique $[p, g, s]$ auprès d'une Autorité de Certification ou CA (Section 2.4.8.2), qui lui fournit alors un cerificat d'authentification. Toutefois, le vendeur garde x secret.

Voici le reste des étapes qui sont exécutées entre le client et le vendeur :

1. Le vendeur et le client doivent avant tout s'assurer qu'il vaut la peine d'exécuter le protocole de négociation. Ils doivent donc comparer le prix maximum que le client peut payer, P^c_{max}, et le prix minimum que le vendeur peut accepter, P^v_{min}, pour le produit sélectionné à la fin de BliS. Pour cela, le vendeur et le client chiffrent conjointement les deux prix et obtiennent respectivement $w_v = E^c_{k''_c}(E^v_{k''_v}(P^v_{min}))$ et $w_c = E^v_{k''_v}(E^c_{k''_c}(P^c_{max}))$. Ils peuvent alors se servir d'une variante (par exemple, [107]) du protocole de Yao pour le *Problème du Millionaire* [184]. À la fin de ce protocole, le client et le vendeur concluent que *la négociation peut avoir lieu* si et seulement si ils sont arrivés au résultat $P^v_{min} \leq P^c_{max}$. Dans le cas contraire, ils abandonnent le processus de négociation.
2. Le vendeur fixe un délai pour la négociation et informe le client de cette contrainte.
3. Si le délai fixé par le vendeur n'est pas encore dépassé :
   (a) Le client choisit aléatoirement $c_1, c_2 \in \mathcal{R}$, formule sa requête masquée $y = (y_1\ ,\ y_2) = (i^2 \cdot g^{c_1}\ ,\ u_v{}^2 \cdot g^{c_2})$, où i désigne l'index du produit choisi et u_v est la valeur d'utilité correspondant à son offre (Section 5.4.1). Il envoie y au vendeur.
   (b) Soit m le nombre de lignes que contient la table M. Le vendeur envoie au client les offres O_ℓ, $\ell = 1, \ldots, m$, qui consistent en des valeurs affectées

aux attributs faisant partie de $\mathcal{A}$ et possiblement la promotion sur le produit choisi. Les offres sont chiffrées comme suit :

$$E_\ell = (\alpha_\ell, \beta_\ell, \gamma_\ell),$$

avec

$$\alpha_\ell = \left((y_1/d_\ell^2)^x\,,\ (y_2/e_\ell^2)^x\right),$$

$$\beta_\ell = \left(g^{v_{1\ell}}\,.\ g^{v_{2\ell}}\right),$$

$$\gamma_\ell = O_\ell^2 \cdot (y_1/d_\ell^2)^{v_{1\ell}} \cdot (y_2/e_\ell^2)^{v_{2\ell}}\,;$$

où $v_{1\ell}$ et $v_{2\ell}$ sont choisis aléatoirement dans $\mathcal{R}$, d_ℓ est l'index du produit dans la ligne ℓ de la table M, et e_ℓ est la valeur d'utilité, pour chaque $\ell = 1, \ldots, m$. Le vendeur envoie E_ℓ, $\ell = 1, \ldots, m$, au client.

(c) Le client a reçu du vendeur m triplets $(\alpha_\ell, \beta_\ell, \gamma_\ell)$. toutefois, les seuls qui sont significatifs pour sa requête sont ceux qui correspondent au couple (d_ℓ, e_ℓ) et qui vérifient $(i, u_v) = (d_\ell, e_\ell)$, s'il en existe (on pourrait avoir aucun ou plusieurs triplets). Pour les localiser, il suffit de remarquer que

$$\alpha_\ell = \left((y_1/d_\ell^2)^x\,,\ (y_2/e_\ell^2)^x\right) = ((i^2 \cdot g^{c_1}/d_\ell^2)^x\,,\ (u_v^2 \cdot g^{c_2}/e_\ell^2)^x),$$

qui est égal à $(g^{c_1 x}, g^{c_2 x}) = (s^{c_1}, s^{c_2})$ si et seulement si $i^2 = d_\ell^2$ et $u_v^2 = e_\ell^2$. Ainsi, le client calcule (s^{c_1}, s^{c_2}) une fois pour toute. Il parcourt par la suite tous les triplets qu'il a reçus du vendeur, à la recherche de ceux qui satisfont $\alpha_\ell = (s^{c_1}, s^{c_2})$. À partir de cet instant, le client peut calculer les offres

$$\gamma_\ell \cdot (\beta_{1\ell}^{c_1} \cdot \beta_{2\ell}^{c_2})^{-1} = O_\ell^2\,,$$

où $\beta_{1\ell}$ et $\beta_{2\ell}$ sont respectivement les première et seconde composantes de β_ℓ. Il obtient O_ℓ en calculant la racine carrée de O_ℓ^2 qui est un élément de $\mathcal{R}$ (Section 2.4.2.2). Si le client est satisfait par l'une des offres, il la sélectionne et poursuit sa procédure d'achat avec le protocole de paiement masqué et de livraison anonyme (BliP). Dans le cas contraire, il peut soit retourner au début de l'étape 3, soit mettre fin au protocole de négociation.

5.5 Paiement masqué et livraison anonyme

Le protocole de paiement masqué et de livraison anonyme (BliP) [8, 11] consiste en deux phases. D'abord, le client doit être en mesure de payer la somme d'argent correspondant au produit qu'il achète sans dévoiler son identité, ses données démographiques ou toute information, en l'occurrence son compte bancaire, qui permettrait de constituer un dossier sur lui. L'argent doit donc être déposé dans le compte bancaire du vendeur sans laisser de traces sur le client.

À ce sujet, nous combinons l'idée de Chaum [56] et le protocole proposé par Aiello, Ishai et Reingold [7] pour construire le protocole de BliP. Avec le protocole proposé dans [56], le client dépose l'argent dans le compte bancaire du vendeur de manière intraçable, en utilisant un pseudonyme dont il se servira plus tard pour acheter des produits auprès du vendeur.

Le protocole décrit dedans [7] concerne la vente des produits numériques au moyen d'un transfert inconscient et d'une forme de débit *pré-autorisé*. Plus précisément, après avoir effectué un dépôt de fonds dans le compte du vendeur, le client est en mesure de mener plusieurs transactions d'achat tant que son compte virtuel ainsi créé est encore créditeur. Le protocole décrit dans [7] permet au vendeur de débiter le compte du client du montant exact de la transaction, même s'il n'est pas en mesure de connaître ledit montant. Ce protocole permet également au vendeur de s'assurer que le solde du compte du client ne deviendra pas débiteur après la transaction.

En cas de nécessité (si les fonds sont insuffisants par exemple), le client peut recharger son compte chez le vendeur, toujours au moyen du protocole de Chaum [56]. De cette façon, il peut se servir de ce compte aussi longtemps qu'il souhaite faire des achats de produits (biens et services) chez le vendeur. S'il décide de fermer son compte, il reçoit du vendeur un chèque électronique anonyme qu'il peut déposer sous un pseudonyme différent dans son compte bancaire principal ou, s'il préfère, dans un compte qu'il pourrait avoir auprès d'un autre vendeur.

Le protocole BliP vise aussi à assurer une livraison anonyme [11, 12] du produit acheté, du vendeur au client. Dans le cas d'un produit numérique, il y a plusieurs possibilités. Si le client est assis dans un café Internet, le produit est téléchargé directement, de préférence sur sa clé USB si la politique du café le permet. Si le client fait confiance à un proxy, il peut s'en servir et recevoir le produit dans le confort de sa maison ou son bureau. Autrement, l'idée d'adresse de retour intraçable proposée par Chaum [54] peut être utilisée.

Le cas d'un produit physique est plutôt complexe car, contrairement à un produit numérique qui circule à travers un réseau électronique, un produit physique exige un expéditeur et un receveur physiques. Dans ce cas, nous avons introduit la notion de centres de livraison anonyme. Nous décrivons dans la section 5.7 l'architecture et le fonctionnement d'un tel centre.

Pour finir, nous signalons qu'après le paiement, le client reçoit un certificat numérique qui lui permettra plus tard (BliM) d'accéder aux services après achat (ou après vente).

5.6 Service après vente masqué

Le masquage du service après vente se fait à travers le protocole BliM (Blind Maintenance) [8]. Il est question de permettre au client de télécharger les dernières mises à jour ou mesures correctives dans une base de données (de service après vente) disponible chez le vendeur. Pour cela, le vendeur doit stocker les dernières mises à jour et/ou les mesures correctives pour ses produits dans une table M. Le protocole BliM est alors semblable à BliS, à la seule différence qu'ici, le client se sert d'un certificat pour interroger la table M afin d'obtenir la mise à jour ou la mesure corrective correspondante. Pour le cas `Showbiz`, Sir Bob pourrait télécharger des instructions qui décrivent de nouvelles méthodes de nettoyage du diamant ou alors, qui indiquent comment s'assurer que le diamant reste bien fixé sur son socle. Pour cela, il doit avoir obtenu d'Alice un certificat à la fin du protocole BliP. Ce certificat est une autre donnée *très sensible* du catalogue d'Alice.

Nous reviendrons, en section 5.7.8, sur le service après vente en termes de gestion de l'après livraison. En fait, nous devons d'abord présenter les centres de livraison anonymes pour mieux justifier cette gestion.

5.7 Centres de livraison anonyme

Dans les étapes précédentes, nous avons montré comment un client peut fouiller de façon masquée dans le catalogue du vendeur, négocier le prix de vente du produit qu'il a choisi, payer de manière intraçable et recevoir la livraion de manière anonyme s'il s'agit d'un produit numérique. La phase de livraison reste toutefois problématique dans la mesure où il est difficile de maintenir l'anonymat d'un client qui achète un produit physique, même si les phases BliS et BliN se sont bien déroulées. En effet, dans le cas d'un produit physique, il se pourrait que le client doive donner son adresse (rue, hôtel, etc.) pour des besoins de livraison. Or, donner cette information permettrait la constitution de dossier sur lui ; ce qui pourrait par la suite fragiliser sa vie privée. Jusqu'ici, la gestion de la vie privée en commerce électronique n'a pas fait l'objet de beaucoup de recherche pour ce qui a trait à la livraison des produits physiques. Tout au plus, l'on a pu noter la présence des tiers de confiance (section 5.7.1). Dans cette section, nous présentons les centres de livraison anonyme ou ADC (*Anonymous Delivery Centres*) [11, 12] dont le rôle est d'assurer la livraison anonyme des produits physiques achetés en commerce électronique.

5.7.1 Généralités sur les ADCs

En général, la livraison est un processus consistant à transférer un paquet d'un expéditeur à un destinataire, en passant éventuellement par un ou plusieurs agents de livraison. Ce processus est illustré dans la figure 5.2.

Un agent de livraison est une compagnie ou un individu qui récupère le paquet à livrer auprès de l'expéditeur ou d'un autre agent de livraison et l'achemine jusqu'au

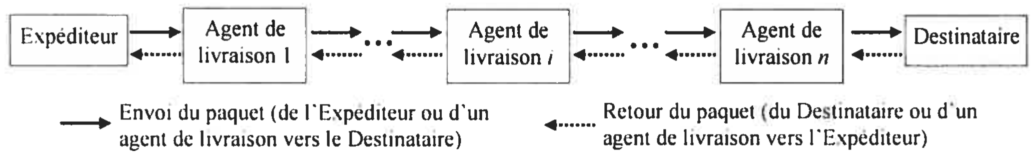


FIG. 5.2 – Entités du processus de livraison.

récepteur ou à un autre agent de livraison. Par exemple, si Alice souhaite envoyer une lettre (physique) à Sir Bob, elle peut le faire par l'entremise d'une compagnie de livraison telle que FedEx, DHL, UPS, etc. S'il s'agit d'un courriel, une ou plusieurs entités sur Internet (ISP ou *Internet Service Provider*, serveurs proxy, etc.) peuvent aider à le convoyer d'Alice à Sir Bob ; chaque entité fait suivre le courriel à la suivante jusqu'à ce que le destinataire final soit atteint.

De manière conventionnelle, le processus de livraison exige des informations d'identification du destinataire et parfois même de l'expéditeur (par exemple, leurs noms et adresses). Cette information permet à l'agent de livraison de trouver soit le prochain agent de livraison à qui le paquet devra être expédié, soit le destinataire final. Pour protéger la vie privée des clients en commerce électronique, la livraison anonyme, tant pour les produits électroniques que physiques, devient un élément indispensable.

Par livraison anonyme, nous entendons le cheminement d'un paquet d'un expéditeur à un destinataire, sans que l'expéditeur soit en mesure d'obtenir les données privées du destinataire. Le paquet dont il est question peut être un courrier, un cadeau, de l'argent, un bien acheté sur Internet, etc. Nous définissons un système de livraison anonyme comme étant un système qui a pour rôle d'assurer la livraison anonyme de paquets, ce qui signifie que l'expéditeur et le destinataire se servent tous du système dans le but de mener à terme le processus de livraison.

Un système de livraison anonyme vise à garantir la vie privée du destinataire tout le long du processus de livraison. La livraison anonyme des produits numériques a été bien étudiée dans le passé. Elle exige que le destinataire—le *client*—soit en mesure de naviguer sur Internet de manière anonyme, ce qui lui permet alors d'envoyer des requêtes et de recevoir des réponses de l'expéditeur —le *vendeur*—(y compris les produits numériques qui sont alors implicitement livrés) sans laisser de trace [7]. Ceci est habituellement réalisé par l'utilisation des tiers de confiance [186, 213]. Si aucun tiers n'est digne de confiance, les *mix-nets* de Chaum [54] peuvent être utilisés et, dans ce cas, une adresse intraçable de retour peut servir pour la livraison du produit numérique.

Les solutions possibles à la livraison anonyme des produits physiques dépendent avant tout du destinataire. De manière globale, on peut dégager deux principaux types de systèmes de livraison anonyme, selon que le destinataire fait ou pas confiance à un tiers.

Si le destinataire fait confiance à un tiers, il donne la procuration à ce dernier pour qu'il récupère le produit auprès de l'expéditeur et le conduise jusqu'à lui. Dans cette catégorie, on peut citer les compagnies *ContinentalRelay* [194] et *Executive Mail Drop Services* [198]. Ces compagnies sont spécialisées dans l'offre de boîtes postales anonymes et de services de transfert de courriers, paquets, cartes postales, messages sur répondeurs, fax et e-mails anonymes. Pour rassurer les individus sur leur vie privée, ces compagnies leur fournissent des services tels qu'une adresse de rue privée, une boîte postale pour ramasser le courrier, ou à une adresse e-mail anonyme. Les paquets qu'elles reçoivent sont par exemple re-adressés et expédiés aux véritables adresses des individus.

Dans un système basé sur un unique tiers de confiance, si celui-ci et l'expéditeur sont de connivence, toute protection de la vie privée du destinataire est perdue. Il est donc utile de penser à une solution qui ne prend pas tout son appui sur un unique tiers de confiance. Un ADC est un système de livraison anonyme dans lequel la vie privée du client ne peut être ménacée que par une coalition d'un nombre significatif d'agents de livraison, de sorte qu'aucune entité prise toute seule n'a besoin d'être digne d'une confiance totale.

5.7.2 Architecture des ADC

Un ADC consiste en trois composantes principales :

- une **unité de dépôt** via laquelle le vendeur ou l'agent de livraison dépose le produit à livrer dans l'ADC,

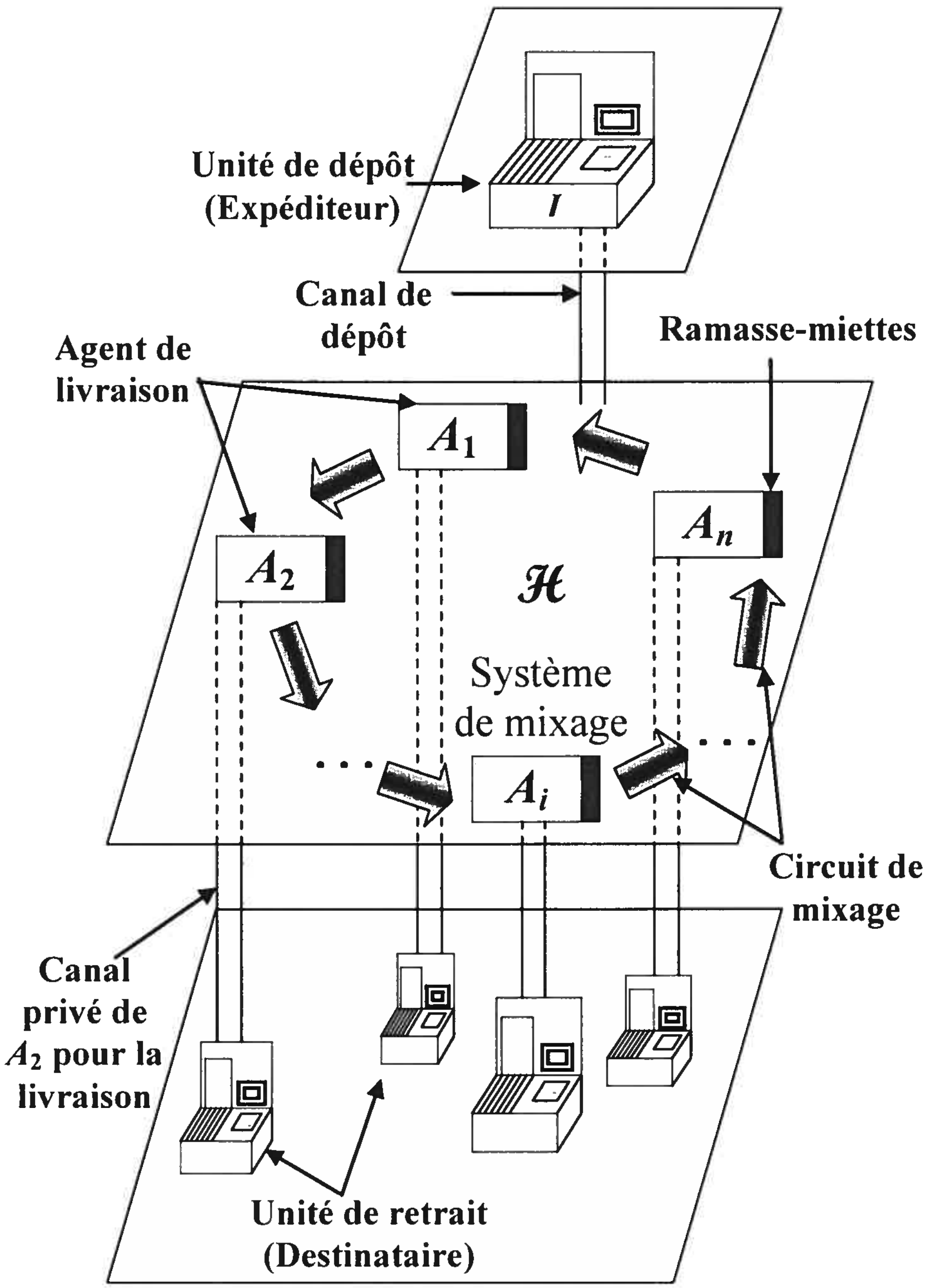


FIG. 5.3 – Centre de livraison anonyme.

- un **système de mixage** qui constitue le processus d'anonymat du produit, qui circule d'un agent de livraison à un autre,
- une **unité de retrait** à travers laquelle le client ou son représentant (un autre agent de livraison par exemple) entre en possession du produit.

Ces trois composantes sont illustrées sur la figure 5.3 et décrites en détail dans les sections 5.7.5 et 5.7.6. Avec l'exemple `Showbiz`, pour livrer la bague à Sir Bob, il se pourrait qu'Alice ait besoin d'informations sur lui (adresse, adresse e-mail, numéro de téléphone, etc.). Mais Sir Bob ne veut lui donner aucune information qui pourrait être utilisée pour entacher sa vie privée. Dès lors, il devrait se servir d'un ADC pour recevoir la livraison de la bague.

5.7.3 Définitions et Notations

Un *Point de livraison* est un endroit où le paquet à livrer subit des transformations, tels que l'emballage et l'étiquetage. L'expéditeur et le destinataire sont des exemples de points de livraison.

Un *Agent de livraison*, A, est une compagnie ou un individu qui transporte le paquet d'un point de livraison à l'autre. En particulier, chaque agent de livraison est en fait un point de livraison. Le système de mixage (figure 5.3) est un réseau constitué de plusieurs agents de livraison.

Un *Centre de livraison anonyme* ou ADC (*Anonymous Delivery Centre*) est un espace physique consacré à la livraison anonyme. L'ADC se compose de plusieurs compartiments qui communiquent entre eux à travers une surface rotative horizontale, $\mathcal{H}$, sur laquelle les paquets sont placés. Chaque compartiment appartient à un agent livraison précis. L'agent de livraison a accès au contenu de $\mathcal{H}$ pour y faire des retraits et des dépôts ; ce qui signifie qu'il peut retirer un paquet se trouvant sur $\mathcal{H}$, lui appliquer des transformations (ex. : re-emballage, re-étiquetage, etc.) avant de le remettre (sur $\mathcal{H}$).

Un *Message mixte*, c, est un message destiné au système de mixage, avec la propriété que chaque agent de livraison a accès au plus à un morceau, m_i, du message tout entier. Le message m_i indique à l'agent de livraison comment manipuler le produit, P qu'il a entre ses mains. Il est composé de deux parties principales : l'action à appliquer sur

P (ex. : re-étiquetage) et le restant du message mixte qu'il faut mettre sur le paquet à l'intention des agents suivants, s'il y a lieu. À tout moment, seule la première partie du message mixte peut être en clair, ce qui signifie alors que seul l'agent de livraison qui détient le produit et qui a été choisi par le client (section 5.7.4) peut déchiffrer cette première partie.

La *Liste des poids*, W, est une liste formée uniquement des poids acceptables dans l'ADC. Lors du dépôt d'un paquet dans l'ADC, l'unité de dépôt choisit au hasard un emballage qui peut contenir le paquet. Elle y met le paquet et remplit l'espace vide restant avec des objets futiles pour aboutir aléatoirement à un poids faisant partie de la liste W. Ce processus porte le nom d'*emballage basé sur le poids*.

Avant de présenter le fonctionnement d'un ADC, nous allons d'abord décrire la constitution d'un message mixte.

5.7.4 Description formelle d'un message mixte

Un message mixte, c, est un texte chiffré formé à partir de t messages (textes en clair ou même textes clairs) $m_1, m_2, \cdots, m_t$ et destiné à des agents de livraison $R_1, R_2, \cdots, R_t \in \{A_1, A_2, \cdots, A_n\}$, choisis par le client (*destinataire*). Le message m_i est adressé à l'agent de livraison R_i. En particulier, R_t est l'agent de livraison *cible* auprès duquel le client viendra retirer le paquet. Le message mixte est fabriqué comme suit :

- Après avoir choisi un ADC, le client y choisit en secret t agents de livraison. Il obtient leurs clés publiques respectives $P_1, P_2, \cdots, P_t$ auprès de l'infrastructure à clé publique de l'ADC[7]
- Le client chiffre le message m_t de l'agent de livraison cible avec la clé publique $P_t : c_t = E_{P_t}(m_t, Stop = d)$, où E est un algorithme de chiffrement publiquement connu, $Stop$ permet de s'assurer que le paquet est entre les mains de l'agent de livraison cible, d est un code dont se servira le client pour récupérer le paquet, et

[7] L'obtention des clés se fait aussi en secret. Pour cela, le client peut par exemple se servir du protocole BliS. Il peut aussi décider de prendre toutes les clés car le nombre d'agents de livraison ne devrait pas être trop grand.

m_t pourrait contenir des instructions additionnelles destinées à R_t.

- Pour $i = t-1$ jusqu'à 1, le client calcule : $c_i = E_{P_i}(m_i, c_{i+1})$, où m_i donne des instructions optionnelles à R_i au sujet du paquet. Le message mixte final est : $c = c_1$.

Nous avons supposé que l'algorithme de chiffrement E n'est pas auto-commutatif, c'est-à-dire que $E_{P_i}(E_{P_j}(m) \neq E_{P_j}(E_{P_i}(m)$ si $i \neq j$, pour un message m donné. Ceci permet de forcer l'ouverture du message mixte dans l'ordre voulu par le client. La procédure d'ouverture fonctionne comme suit :

- Pour $i = 1$ à t, l'agent de livraison R_i calcule : $D_{K_i}(c_i) = D_{K_i}(E_{P_i}(m_i, c_{i+1})) = (m_i, c_{i+1})$, où K_i est la clé privée correspondant à la clé publique P_i. Il obtient le texte en clair m_i, ainsi que le texte chiffré c_{i+1}. Le texte en clair m_i indique les actions que R_i doit appliquer sur le paquet (ex. : ajouter un emballage par-dessus), tandis que c_{i+1} est la partie restante du message mixte que R_i met sur le paquet au profit de R_{i+1}, pour $i < t$. Il est important de noter que R_i accomplit cette tâche sans avoir besoin de savoir qui est R_{i+1}. En fin de compte, l'agent de livraison cible, R_t, reçoit le message $(m_t, Stop = d)$, garde le paquet, mémorise d et attend le client pour le retrait.

Les agents de livraison $R_1, R_2, \ldots, R_t$ sont choisis par le client à partir de l'ensemble de tous les agents de livraison $\{A_1, A_2, \ldots, A_n\}$ de l'ADC, avec $t < n$. Entre R_i et R_{i+1}, il peut y avoir d'autres agents $A_{i_1}, A_{i_2}, \ldots, A_{i_s}$ de sorte que l'agent A_{i_1} reçoit le paquet de R_i, A_{i_2} le reçoit de $A_{i_1}, \ldots$, et R_{i+1} le reçoit de A_{i_s}. Toutefois, rien dans le message mixte n'indique explicitement qui est le prochain agent auquel il est destiné. Par conséquent, tous les agents intervenants, $A_{i_1}, A_{i_2} \ldots, A_{i_s}$, utilisent leur clé secrète pour essayer de déchiffrer le message mixte. Si la première partie du message est inintelligible, ils laissent simplement le paquet poursuivre son chemin. C'est ici une différence essentielle entre les messages mélangés utilisés dans les ADC et la technique de mix-nets de Chaum [54], dans laquelle chaque intervenant apprend l'identité de son suivant.

5.7.5 Fonctionnement d'un ADC

Nous avons introduit la notion d'ADC pour faciliter la livraison anonyme des produits physiques, en se servant de plusieurs tiers de confiance, appelés Agents de livraison. Nous allons illustrer le fonctionnement d'un ADC avec l'exemple Showbiz.

Après la phase de paiement, Sir Bob choisit un ADC ainsi que t agents de livraison. Il choisit des actions éventuelles, m_i, pour chaque agent R_i et calcule le message mixte correspondant, c, tel qu'expliqué en section 5.7.4. Puis, Sir Bob annonce à Alice l'ADC qu'il a choisi et lui donne c. Alice (ou son agent de livraison) se présente avec le paquet (contenant la bague) auprès de l'unité de dépôt, U, de l'ADC (figure 5.3) et lui remet le message mixte de Sir Bob : c. Le rôle principal de U est d'appliquer la procédure d'emballage basé sur le poids expliquée en section 5.7.3 et de délivrer un reçu à Alice (ou à son agent de livraison) comme preuve de dépôt.

Le système de mixage prend alors le contrôle du paquet. En se rappelant que chaque agent de livraison d'un système de mixage est en fait un point de livraison, considérons l'ensemble de tous les agents de livraison de l'ADC $S = \{A_1, A_2, \ldots, A_n\}$. Nous supposons que chaque agent de livraison A_i prenne exactement le même temps pour agir sur n'importe quel paquet circulant sur $\mathcal{H}$. Sans perte de généralité, nous considérons que $\mathcal{H}$ tourne autour des points de livraison selon leur ordre dans S. Ainsi, A_1 est le premier qui reçoit le paquet provenant de l'unité de dépôt U. A_1 essaie donc de déchiffrer le message mixte de Sir Bob, en utilisant sa clé secrète. S'il réussit, il applique l'action contenue dans le message, le cas échéant, et ré-étiquette le paquet avec la partie restante du message mixte, tel qu'expliqué dans la section 5.7.4. Si le déchiffrement du message mixte échoue (parce qu'A_1 n'était pas le premier agent de livraison choisi par Sir Bob), alors A_1 remet simplement le paquet sur $\mathcal{H}$. Le processus continue, d'un agent de livraison à un autre, jusqu'à ce que le paquet atteigne l'agent de livraison cible, R_t, choisi par Sir Bob. R_t applique également toute action prescrite par Sir Bob dans le message, m_t, à lui, adressé. Il garde le paquet et introduit un paquet factice dans $\mathcal{H}$, indistinguable du paquet qu'il a gardé, de sorte qu'aucun agent ne se rend compte que c'est lui l'agent cible. Les paquets factices ainsi introduits dans $\mathcal{H}$ peuvent être identifiés après un certain temps à cause de leurs étiquettes qui ne changent pas pendant un tour complet sur $\mathcal{H}$. Toutefois, ceci ne compromet pas l'identité de l'agent cible. Ces paquets factices sont rétirés de $\mathcal{H}$ après un intervalle de temps régulier par un *ramasse-miettes* et sont recyclés pour un usage ultérieur.

Pour faire le retrait de son paquet, Sir Bob connaît son agent de livraison cible R_t, puisque c'est lui qui l'a choisi, ainsi que le code, $Stop = d$, qu'il a inséré dans le message destiné à R_t. En fait, $d = f(x)$, pour f une fonction à sens unique publiquement connue et une entrée secrète x connue seulement de Sir Bob. Afin de convaincre R_t que le paquet lui appartient, Sir Bob doit prouver sa connaissance de x. Selon le niveau de la confiance qu'il a en R_t, il peut décider de faire une preuve à divulgation nulle [37, 81] afin que R_t ne prenne pas du tout connaissance de x. Aussi, au lieu de retirer lui même le paquet, Sir Bob pourrait envoyer un autre agent de livraison A' pour récupérer le paquet et le transférer à un autre ADC, et le processus entier recommence (d'une nouvelle unité de dépôt U, à un nouvel agent cible R_t). Pour ce faire, Sir Bob doit aider A' à convaincre R_t de sa connaissance de x.

5.7.6 Description formelle d'un centre de livraison anonyme

Les composantes de l'ADC fonctionnent au moyen de trois algorithmes dont le pseudo-code est donné ci-dessous.

AlgorithmeDépôt(c)

1. Demander au vendeur de fournir le message mixte c.
2. Ouvrir la porte de l'unité de dépôt et initialiser un *délai* de terminaison.
3. Demander au vendeur d'introduire le paquet dans l'unité de dépôt.
4. Fermer la porte de l'unité de dépôt dès que le paquet y est déposé ou si le délai est expiré.
   (a) Si un paquet a été déposé, appliquer la procédure d'emballage basé sur le poids et étiqueter le paquet avec le message mixte. Transférer le paquet

marqué au système de mixage. Délivrer un reçu au vendeur contenant, entre autres, le code c, la date de dépôt et des informations sur l'ADC.

(b) Si le délai est expiré, annuler l'opération et demander au vendeur d'essayer plus tard.

AlgorithmeMixage(c)

1. Le paquet, P, étiquetté c, va d'un agent de livraison à l'autre. Aussi longtemps que P circule dans le système de mixage, chaque agent de livraison A le récupère et essaie de déchiffrer l'étiquette c en utilisant sa clé secrète K. En cas d'échec, P est remis sur le système à l'intention du prochain agent. En cas de succès, on a : $D_K(c) = (m, c')$. L'agent A réalise la tâche m s'il y en a.
   (a) Si c' est de la forme $Stop = d$, c'est que l'agent cible R_t a été atteint ; il garde le paquet, mémorise d, et fait suivre à un paquet factice, indistinguable de P, au prochain agent ;
   (b) Sinon l'agent A re-étiquette le paquet avec c' et le laisse aller au prochain agent.
2. L'agent cible R_t se met en attente du client pour le retrait du paquet—exécution de *AlgorithmeRetrait*(d).

AlgorithmeRetrait(d)

1. Le client (ou son représentant) accède à l'unité de retrait de son agent de livraison cible R_t.
2. R_t demande au client de prouver sa connaissance de x tel que $f(x) = d$.
   (a) Si la vérification échoue, annuler l'opération et demander au client d'essayer plus tard.
   (b) Si la vérification se fait avec succès, alors R_t transfère le paquet à l'unité de retrait, ouvre la porte de cette unité et initialise un délai.
   (c) R_t demande à l'utilisateur de retirer le paquet.

(d) R_t ferme la porte de l'unité de retrait une fois que le paquet a été retiré ou alors que le délai est expiré.

(e) Si le délai est expiré, annuler l'opération et demander au client d'essayer plus tard.

5.7.7 Illustration avec *Showbiz*

Nous allons simuler l'utilisation d'un ADC dans le cadre de Showbiz. Sir Bob choisit un ADC pour la livraison de la bague de fiançailles de Claudia. Supposons pour simplifier que l'ADC se compose de cinq agents A_1, A_2, A_3, A_4, A_5. Sir Bob choisit trois de ces agents, disons $R_1 = A_2$, $R_2 = A_4$ et $R_3 = A_3$. Il accède à leurs clés publiques respectives P_2, P_4, P_3 et s'en sert pour créer le message mixte, $c = E_{P_2}(m_1, E_{P_4}(m_2, E_{P_3}(m_3, Stop = d)))$, à partir du texte en clair $m = (m_1, m_2, m_3)$, où m_1 = "SVP, faire suivre", m_2 = "SVP, retenir le paquet pendant un jour", m_3 = "SVP, emballer le paquet dans un sac brun", et $d = f(x)$ pour un certain x qu'il a choisi en secret. Les messages m_i sont fabriqués par le client lui-même, selon ses desiderata et ses contraintes. Par exemple, le contenu du message m_2 dans le cas de Showbiz peut signifier que Sir Bob accorde peu d'importance au délai de livraison.

Sir Bob annonce par la suite à Alice l'ADC qu'il a choisi et lui envoie c. Alice emballe la bague et se présente à l'unité de dépôt avec le paquet. Elle se sert de *AlgorithmeDépôt*(c) pour introduire le paquet dans l'ADC.

L'unité de dépôt applique au paquet la procédure d'emballage basé sur le poids et met le paquet sur $\mathcal{H}$, à l'intention du point de livraison A_1. L'agent de livraison A_1 ne change rien sur le paquet puisque Sir Bob ne l'a pas choisi lors de la fabrication du message mixte. Plus précisément, A_1 essaie de déchiffrer c, découvre du baragouin et, par conséquent, il laisse le paquet poursuivre son chemin (vers A_2).

Le prochain agent de livraison, A_2, calcule :

$$D_{K_2}(c) = D_{K_2}(E_{P_2}(m_1, E_{P_3}(m_2, E_{P_1}(m_3, Stop = d))))$$

et obtient le message m_1, ainsi que le texte chiffré $c_1 = E_{P_3}(m_2, E_{P_1}(m_3, Stop = d))$.

Puisque m_1 est un message intelligible, A_2 re-étiquette simplement le paquet avec le nouveau message mixte c_1 et fait suivre le paquet (comme indiqué dans m_1), en le remettant sur $\mathcal{H}$.

L'agent de livraison A_3 ne change rien sur le paquet pour l'instant car, malgré le fait qu'il ait été choisi par Sir Bob, il ne le saura qu'après le re-étiquetage du paquet par A_4. En fait, le déchiffrement du message mixte c_1 ne donne rien d'intelligible lors du premier passage du paquet par A_3. Ce dernier laisse donc le paquet poursuivre son chemin (vers A_4).

A_4 déchiffre c_1 et obtient m_2 et c_2. Comme le stipule le contenu de m_2, A_4 garde le paquet pendant un jour avant de le remettre sur $\mathcal{H}$. Pour voiler cela, d'une part, A_3 insère en ce moment un paquet factice dans $\mathcal{H}$ et, d'autre part, avant de remettre le paquet sur $\mathcal{H}$, il recommence par l'insertion d'un autre paquet factice qu'il remplace après un tour complet par le véritable paquet.

Le jour suivant, A_3 reçoit le paquet, déchiffre c_2 et découvre qu'il a été choisi comme agent cible du paquet qu'il a en main. A_3 met la bague dans un sac brun (bien sûr, sans qu'il sache qu'il s'agit d'une bague, puisque l'emballage original provenant de l'unité de dépôt est toujours en place). Plus tard, Sir Bob se présente à l'unité de livraison de A_3 et se sert de *AlgorithmeRetrait*(*d*), et de sa connaissance de x pour retirer le paquet. Il pourrait également demander à un nouvel agent de livraison, A', de retirer le paquet à sa place et de le transférer à un autre ADC, où tout le processus devrait recommencer. Dans ce dernier cas, Sir Bob pourrait révéler x à A'.

5.7.8 Gestion de l'après livraison

Parfois, il est souhaitable que le client soit rejoint après avoir reçu la livraison d'un paquet, par exemple s'il s'avère que le produit acheté par le client est défectueux ou même dangereux. Il faudrait alors permettre à l'ADC de pouvoir rejoindre le client. Une solution à ce problème consiste à se servir des adresses de retour de courriel intraçables, telles que proposées par Chaum [54]. Le client pourrait aussi interroger de manière masquée la base de données du service après vente du client sur une base régulière, en utilisant BliS. De plus, si le client n'est pas satisfait du produit (paquet) qu'il a reçu (de manière anonyme), il doit être en mesure de le retourner au vendeur, en se servant éventuellement d'un autre ADC qui, dans ce cas, doit être capable de cacher l'identité de l'expéditeur (le client).

5.8 Systèmes de recommandation préservant la vie privée

Comme nous l'avons dit dans la section 1.6, dans un contexte du commerce électronique, les Systèmes de Recommendation (SR) permettent au vendeur d'assister ses clients dans la recherche de produits pouvant satisfaire leurs besoins. Grâce aux SR, les clients n'ont pas besoin de consulter le catalogue du vendeur dans son intégralité car le vendeur se sert des techniques de filtrage pour proposer efficacement des produits qui pourraient intéresser le client. Pour cela, le vendeur a généralement besoin d'avoir de l'information sur le client. Comme nous l'avons vu au chapitre 3, cela revient à constituer des dossiers sur les clients, élément que nous combattons dans le cadre de cette thèse, étant entendu que toute constitution de dossiers peut être considérée comme point de départ d'une violation de vie privée. Dans cette section, nous présentons le système de recommendation ALAMBIC [13, 14], que nous avons conçu dans une perspective de préservation de la vie privée.

ALAMBIC est un SR hybride qui combine trois techniques de filtrage, à savoir : le filtrage par contenu, le filtrage démographique et le filtrage collaboratif. Il est basé sur l'utilisation d'un tiers de *semi-confiance*, c'est-à-dire que le client ne fait confiance à ce tiers que pour une partie de ses données privées. Plus précisément, le client confie une partie de ses données au tiers et l'autre au vendeur, de telle sorte que les deux sont en mesure de lui recommander des produits, mais sans que l'un ou l'autre puisse tout seul déduire de l'information qui pourrait compromettre sa vie privée. En d'autres termes, la vie privée du client est préservée aussi longtemps que le tiers de semi-confiance et le vendeur ne forment pas une coalition.

Dans la section 1.6, nous avons présenté les cinq principaux SR tels que vus par Burke [44]. Dans le cadre du système ALAMBIC, nous nous sommes limité aux trois techniques que nous venons de mentionner car ce sont elles qui sollicitent davantage de données privées sur des clients. En effet, dans un contexte de filtrage collaboratif et de filtrage par contenu, le client fournit typiquement au vendeur de l'information sur son comportement d'achats, en particulier sur les produits qu'il a achetés dans le passé, ainsi que ses votes sur ces produits. Cela est dû au fait que le calcul de prédictions et de similarité est basé sur une matrice qui stocke les votes que des clients ont fait sur des produits. Pour ce qui est du filtrage démographique, les clients fournissent de l'information démographique au vendeur afin de permettre leur catégorisation.

Pour décrire notre système ALAMBIC, d'une part (section 5.8.1), nous en présentons les concepts et principes généraux, ainsi que l'architecture globale et ses composants. D'autre part (section 5.8.2), nous donnons un détail des différentes procédures de filtrage.

5.8.1 Architecture et composantes

ALAMBIC est un système qui offre une architecture et des protocole génériques en vue de développer des systèmes de recommandation basés simlutanément sur les techniques de filtrage démographique, collaboratif et par contenu. C'est un système qui donne de l'assurance pour ce qui est (1) de la préservation de la vie privée du client et, (2) de la protection des données sensibles du vendeur. Pour ce faire, ALAMBIC utilise, en plus des primitives cryptographiques bien connues, le principe suivant :

> "Trust no one, but you may trust two..."

En d'autres termes, le client pourrait faire confiance à deux entités même s'il n'a confiance à aucune d'elles prise individuellement. Le client pourrait donc ne pas faire confiance au vendeur pour ses données démographiques, son comportement d'achats dans le passé, etc., mais il pourrait confier ces informations à une autre entité de bonne réputation. Toutefois, même si le client fait confiance à une telle entité pour ses données privées, il pourrait ne pas vouloir qu'elle sache quoi que ce soit sur ses achats, les produits qui l'intéressent ou, d'une manière générale, le genre de transactions qu'il mène avec le vendeur.

Le système ALAMBIC utilise ce principe pour permettre au client de partager sa confiance entre le vendeur et un *tiers de semi-confiance* indépendant, appelé *still maker*, de telle manière que seule une connivence entre ces deux entités (vendeur et still maker) peut compromettre la vie privée du client. La confiance partielle exigée du still maker n'est pas une invention en soi. Elle est semblable à la confiance traditionnellement mise en tous ceux-là qui sont liés par le sceau de la confidentialité (ex. : notaires, autorités de certification, etc.) et dont les affaires dépendent en grande partie de la confiance du public, de leur honnêteté et de leur professionnalisme.

Le still maker dispose d'une plateforme, appelée STILLMAKER, à travers laquelle il produit des agents appelés, *agents Alambic*, et destinés à être déployés dans les plateformes des vendeurs, pour accomplir les tâches nécessaires aux processus de recommandation. Un agent Alambic s'exécute directement dans la plateforme du vendeur et n'a plus de contact subséquent avec son créateur, le still maker.

Nous allons maintenant donner plus de détails sur les diverses entités en présence ainsi que les composantes de l'architecture d'ALAMBIC. Cette architecture est présentée dans la figure 5.4.

La plateforme du vendeur : C'est ici que réside le catalogue des produits disponibles chez le vendeur. Cette plateforme garde également une base de données des profils chiffrés des clients, ainsi qu'une base de données contenant les votes chiffrés des clients. Malgré le fait que ces données sont gardées dans la plateforme du vendeur, elles ne peuvent pas être interprétées ou utilisées sans la collaboration de l'agent Alambic. La plateforme du vendeur communique avec l'agent Alambic et le client par l'intermédiaire d'une unité de contrôle.

Le STILLMAKER : C'est une plateforme placée sous la seule autorité du still maker. Elle génère un agent Alambic distinct pour chaque plateforme de vendeur à la demande de ce dernier. Chaque agent Alambic dispose de sa paire de clés publique/privée et d'un certificat à clé publique signé par le still maker, à travers une infrastructure à clé publique (PKI) connue à la fois du vendeur et du client.

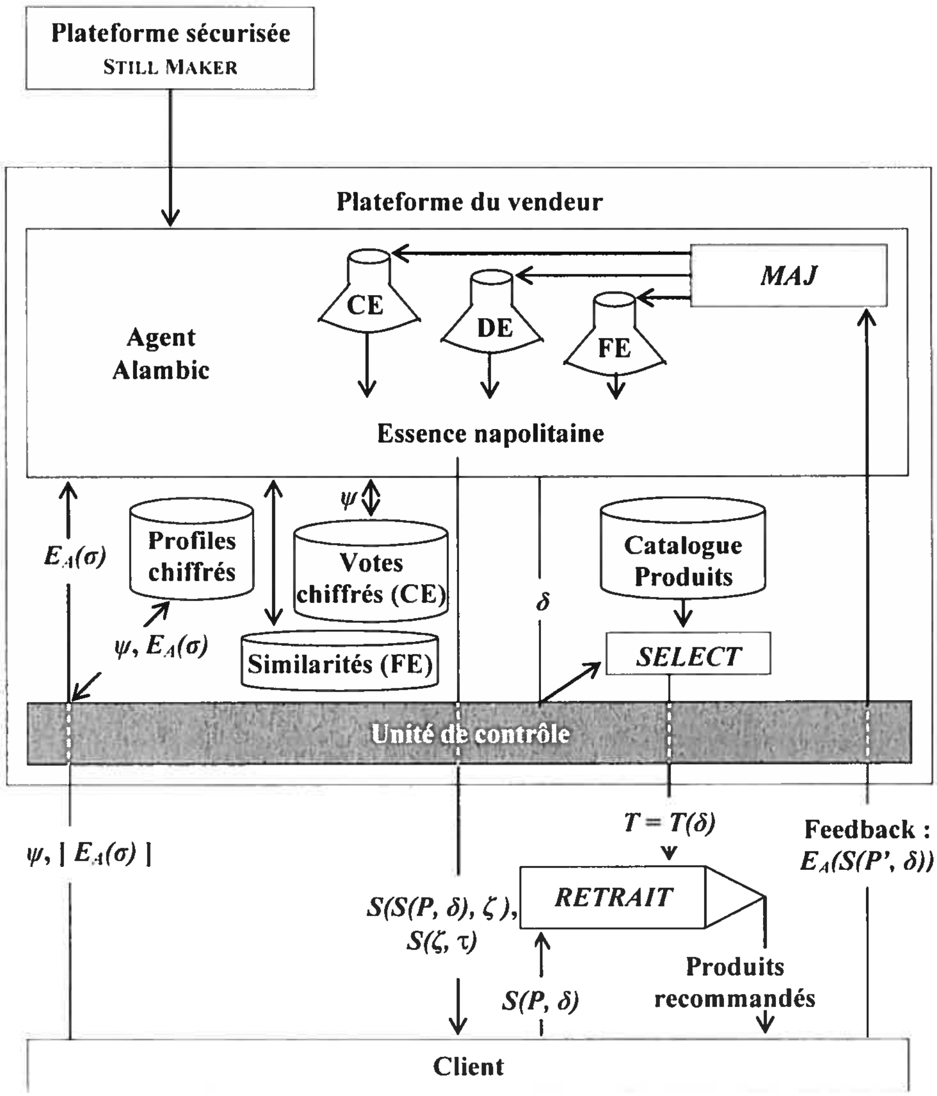


FIG. 5.4 – Architecture du système ALAMBIC. Les interactions entre les différentes composantes sont illustrées de gauche à droite.

Le client : Il souhaite recevoir des recommandations de produits de la part du vendeur. Il obtient auprès du vendeur le certificat à clé publique de l'agent Alambic et peut vérifier son authenticité puisqu'il a indépendamment obtenu la clé publique du still maker auprès de la PKI. Il peut dès lors chiffrer ses données privées (données démographiques et votes sur les produits) avec la clé publique de l'agent Alambic afin de les protéger du vendeur.

L'agent Alambic : Cet agent sert d'intermédiaire entre le vendeur et le client. C'est lui qui gère les statistiques sur les données démographiques et le feedback reçus des clients, afin de produire des recommandations. L'agent d'Alambic peut, par exemple, être un logiciel que le vendeur télécharge du STILLMAKER et installe dans son système. Dans le souci d'assurer l'intégrité de la plateforme du vendeur, le code de l'agent Alambic est numériquement signé par le still maker. Inversement, dans le souci d'assurer l'intégrité de l'agent Alambic et, par conséquent, l'intégrité de n'importe quelles données qu'il maintient, son code est chiffré et obfusqué (section 5.2). De cette façon, le vendeur ne peut obtenir plus d'informations en l'exécutant sur sa plateforme que ce qui peut être déduit des entrées et des sorties des divers processus dans lesquels l'agent d'Alambic est impliqué. Aussi, l'agent Alambic peut être fourni sous forme d'un dispositif physique tels qu'une *carte à puce*, un *co-processeur* sécurisé, etc., que le vendeur ajoute à son système.

5.8.2 Procédures de filtrage

L'exemple suivant illustre pourquoi nous avons besoin des procédures de filtrage en général, et des procédures de filtrage qui préservent la vie privée en particulier, dans un contexte de commerce électronique. Un autre exemple est donné en annexe I sous la forme d'une conversation...

Exemple - Énoncé 1.

De nos jours, un citoyen qui veut être bien informé sur les affaires et la politique du monde fait face à une myriade de sources d'informations. Même si nous le limitons aux *sources d'Internet, il y a des centaines de sources parmi lesquelles on retrouve de simples médias, des sites Web de parties politiques, des groupes de lobbying et d'intérêt, ainsi que des newsgroups et autres blogs*[8]*. Au-delà du nombre d'orientations de telles sources d'informations sur un sujet controversé (par exemple, la guerre en Irak), un utilisateur pourrait avoir des préférences sur la source à employer. Ces préférences peuvent être basées sur la longueur et la profondeur des analyses faites, l'orientation politique, l'objectivité, les moyens de livraison, etc. En ce qui nous concerne, nous allons considérer qu'un certain vendeur dispose d'un moteur de recherche permettant de trouver des articles relatifs aux affaires et à la politique dans le monde. Ses clients sont alors des individus à la recherche des articles à lire. Les liens (URL ou* `Uniform Resource Locator`*) retournés en réponse à une requête initiale*[9] *de l'usager vont constituer le catalogue des produits (ou des articles). Ces liens sont alors sensés répondre ne serait-ce que de façon minimale aux besoins du client (le citoyen). Il est donc intéressant de se servir d'un système de recommendation afin de présenter au client, de manière prioritaire, les liens pertinents*[10] *qui pourraient satisfaire ses intérêts. Des données implicites sur les préférences et le comportement de l'utilisateur peuvent être recueillies par le système en gardant la trace des liens (URL) qui ont été visités à partir du moteur de recherche. Une telle trace porte le nom de* `clikstream`*.*

Comme nous l'avons déjà dit, notre architecture permet d'implémenter simultanément trois types de filtrage (démographique, collaboratif et par contenu). Pour cela, l'agent Alambic gère séparément les données requises pour chacune des techniques pour faire des recommandations. Voyons donc comment intervient chacune des techniques dans l'énoncé qui précède.

Puisqu'il est raisonnable de penser que des préférences de l'utilisateur seront influencées par des caractéristiques démographiques tels que l'âge, le genre, le niveau de l'éducation, la richesse, l'endroit géographique, etc., nous présumons que le moteur de recherche emploie un système de recommandation basé sur le `filtrage`

[8] *Ce sont des journaux personnels qu'on rend disponibles sur Internet.*
[9] *Recherche à partir des mots-clés par exemple.*
[10] *Il se pourrait que le nombre de tels liens soit grand.*

démographique pour catégoriser les liens visités par un utilisateur donné. Un tel système de recommandation va alors produire des recommandations de la forme : "Les utilisateurs ayant des caractéristiques semblables aux vôtres ont prioritairement visité les adresses URL (liens) suivantes..."

Aussi, un système de recommandation basé sur le filtrage collaboratif se servirait des corrélations entre les clickstreams pour trouver les utilisateurs précédents avec les clicktreams semblables et faire des recommandations basées sur les liens qu'ils ont suivis. Une recommandation pourrait par exemple avoir la forme : "les utilisateurs qui sont passés par l'URL que vous venez de visiter ont aussi visité les URL suivantes...".

Finalement, un système de recommandation basé sur le contenu passerait par la liste entière de liens et calculerait des valeurs de similarité entre ces liens et ceux que le même utilisateur a suivis dans le passé. Les valeurs de similarité seraient basées sur un vecteur de valeurs des caractéristiques tels que les moyens de livraison (fichier texte, vidéo, audio, etc.), l'âge de l'article, la langue, le type de la source (commercial, indépendant, gouvernemental, etc.), la longueur, etc. Les recommandations seraient extraites à partir de la nouvelle liste de liens en choisissant ceux qui sont les plus semblables à ceux choisis dans le passé par l'utilisateur.

Voici du reste la description des trois techniques utilisées : Dans le cas du *filtrage démographique*, l'information exigée est la liste des clusters de clients, accompagnée d'une correspondance entre ces clusters et les agrégats des préférences relatifs aux clients dans ces clusters, par leurs comportements et feedback issus des achats précédents. Nous désignons ces agrégats de préférences par **essence démographique** ou DE ; ce qui signifie que l'agent Alambic distille cette essence à partir des données brutes qu'il obtient lors de ses interactions avec des clients. Afin de protéger les intérêts du vendeur, il faut empêcher toute utilisation non autorisée par l'agent Alambic. Pour cela, la liste de produits évalués par les clients d'un cluster est masquée par le truchement des index anonymes p_i, qui sont liés à l'index réel du produit s_i, à travers une correspondance Σ connue seulement du vendeur. La structure de DE est illustrée dans le Tableau 5.4.

Dans le cas du *filtrage collaboratif*, il y a trois points à considérer :

- *Filtrage collaboratif basé sur la mémoire* : L'agent Alambic gère une matrice des

TAB. 5.4 – Tableau de catégorisation de l'agent Alambic (essence démographique).

Cluster	Centroïde	Index anonymes
C_1	(x_1, y_1, z_1)	$p_{d,1}, p_{d,2}, p_{d,7}$
C_2	(x_2, y_2, z_2)	$p_{d,5}, p_{d,8}, p_{d,20}, p_{d,t}$
...	...	...
C_n	(x_n, y_n, z_n)	$p_{d,10}$

Remarque : À chaque cluster de clients, C_j, correspondent un centroïde avec des coordonnées et une liste d'index anonymes des produits achetés ou sélectionnés par les clients issus de C_j.

votes sur les produits. La distillation d'essence est alors basée sur cette matrice. (C'est ce cas que nous avons illustré pour des besoins de simplicité).

- *Filtrage collaboratif basé sur un modèle* : L'agent Alambic gère une liste de clusters de clients en fonction de leur similarité dans les votes (ceci est semblable au cas des clusters de DE). Il peut aussi gérer la description du modèle en s'appuyant sur un réseau bayésien pour calculer les recommandations.
- *Filtrage collaboratif hybride (mémoire et modèle)* : Ceci exige simplement que le réseau bayésien prenne appui sur la matrice des votes.

Dans tous les cas, nous parlons d'*essence collaborative* ou CE (*Collaborative Essence*), pour référer à l'information gardée par l'agent Alambic en vue de faire du filtrage collaborative.

Pour ce qui du *filtrage par contenu*, l'agent Alambic doit garder une matrice des similarités, $Sim(i, j)$, calculées à partir de l'équation 1.3 (chapitre 1). Afin de protéger les données du vendeur, ces valeurs de similarité devraient être calculées par celui-ci et rendues disponibles à l'agent Alambic. Ceci empêche l'agent Alambic de devoir connaître les caractéristiques de chaque produit. De plus, le vendeur est en mesure de mettre cette essence à jour en tenant compte de l'évolution de son catalogue de produits ; il le fait sans risque de violation de la vie privée pour ses clients puisque l'essence ne contient pas de données sur eux. En outre, l'agent Alambic doit avoir accès aux votes du client courant afin de calculer des prédictions de produits conformément à l'équation 1.4. Il faut cependant noter que l'agent Alambic n'a pas besoin d'avoir accès à la matrice des votes toute entière, mais seulement à la ligne de la matrice qui correspond au client courant.

En d'autres termes, la matrice $Sim(i, j)$ est la seule chose nécessaire à la constitution de l'*essence basée sur le contenu* que nous appelons FE (*Feature Essence*).

L'énoncé 2 (ci-dessous) de notre exemple met en relief le besoin de créer des agrégats (clusters) d'utilisateurs, de disposer d'une historique de ceux qui ont visité le système par le passé, ainsi que la prise en compte du contenu des liens.

Exemple - Enoncé 2. *Supposons que notre citoyen soit un jeune homme étudiant en ingénierie vivant dans un pays européen nordique, d'origine arabe, qui aime le football, conduit une voiture jaune, et a indiqué qu'il est très intéressé par les affaires et les jeux sur ordinateur. Ces caractéristiques, entre autres, constituent son profil démographique. Le système de recommandation basé sur le filtrage démographique l'a identifié comme faisant partie d'un cluster des bonshommes bien instruits des pays développés. Historiquement, les gens dans ce cluster ont la plupart du temps cliquer sur des liens menant à des magazines spécialisés, à un certain nombre de bloggers prolifiques, et aux clips multimédia des radiodiffuseurs d'Europe occidentale. Quand il interroge le moteur de recherche pour des articles au sujet de la guerre en Irak la semaine dernière, il reçoit, dans l'ordre, un article détaillé du journal The Economist, deux* `posts`[11] *des bloggers au sujet du dernier bombardement, ainsi qu'un clip audio de Deutsche Welle. Bien sûr, il y a d'autres liens sur le sujet, mais ils apparaissent à la suite de ceux que nous venons de donner.*

D'autre part, le système de filtrage collaboratif aurait pu identifier que son clickstream suit toujours un certain modèle (par exemple, les sessions incluent normalement deux ou trois clics sur des liens vers les sites Web de Al-Jazeera, CNN et Deutsche Welle en anglais). Par conséquent, et étant donné l'historique de la plupart des personnes ayant visité ces sites, le système de recommandation recommandera des clips vidéo médias de BBC sur le même sujet (guerre en Irak).

Aussi, en utilisant les méthodes d'extraction de caractéristiques basées sur des mots-clés dans le texte des pages Web, le système de recommandation basé sur le contenu trouve des articles similaires à ceux dont les liens ont reçu un clic de la part de notre

[11] *Articles écrits dans un blog.*

jeune citoyen par le passé, par exemple, ceux issus des mots-clés "football", "bière" au format "vidéo", etc.

En réalité, pour décrire les procédures de filtrage, nous avons besoin de choisir le niveau de tolérance que le client veut appliquer à ses données privées. Nous avons choisi la `tolérance zéro` (section 4.2.1), car c'est celle qui exige la protection de toutes les données de la SLP (Stop List for Privacy) du client.

Il est également utile de sélectionner le niveau de tolérance au tracking. En principe, le niveau idéal est *tolérance zéro au tracking*, c'est-à-dire que le client ne veut pas qu'on garde des traces sur ses passages par le système. Considérons le scénario suivant :

Exemple - Enoncé 3.

Notre jeune étudiant pourrait ne pas vouloir qu'on sache qu'il parle arabe (c'est un élément de son profil démographique). Cependant, cela pourrait être approprié pour que le système de recommandation suggère des articles provenant de Al-Jazeera. En même temps, ses précédents clickstreams pourraient inclure des articles avec une moralité douteuse qu'il a beaucoup aimés et, en conséquence, il les as évalués positivement. Il tient certainement à ce que ce qui précède soit gardé confidentiel. En outre, dans un scénario de tolérance zéro aux données privées, il ne voudrait pas que le vendeur (propriétaire du moteur de recherche), ou même quelqu'un d'autre, sache qu'aujourd'hui, il recherche des articles sur des événements récents en Irak. Enfin, alors qu'il avait l'habitude d'utiliser un compte anonyme ouvert dans `Yahoo.com` *pour s'identifier au système de recommandation du vendeur, il pourrait être inquiet que celui-ci sache d'une connexion à une autre qu'il est la même personne. Il préfère ainsi saisir ses préférences de manière anonyme chaque fois qu'il utilise le système plutôt que de faire savoir au vendeur qu'il est ce même type qui s'est servi du système et qui a surfé jusqu'à 4 h 03 mn du matin il y a de cela cinq jours. Ceci justifie donc son choix d'un niveau de tolérance zéro au tracking.*

D'autre part, supposons que le modèle d'affaires du vendeur consiste à charger un certain montant d'argent en fonction du nombre de liens retournés. Par exemple, on peut supposer que le vendeur demande 1$ pour 25 requêtes devant retourner jusqu'à 3 liens chacune. Les liens additionnels sont disponibles à des coûts supplémentaires. En *vue de satisfaire ses clients, le vendeur devrait utiliser un système de recommandation pour trouver les 3 meilleurs liens à retourner à l'utilisateur. En outre, il ne veut pas laisser l'utilisateur passer en revue tous les liens, parce que cela engendrerait un manque à gagner (après tout, faire fonctionner un bon moteur de recherche devrait rapporter de l'argent !). Finalement, savoir quels liens fournir à quels utilisateurs (l'essence) est l'information sur laquelle le modèle d'affaires tout entier est basé. Perdre cette information au profit d'un autre vendeur (moteur de recherche) renverrait à une perte de son avantage sur la concurrence.*

Malgré ce qui précède, dans le souci de simplifier l'écriture des procédures de filtrage, nous avons considéré une `tolérance totale au tracking`[12] de la part du client. En particulier, toutes les procédures de filtrage écrites dans la présente section sont basées sur les hypothèses "tolérance zéro sur les données privées" et "tolérance totale au tracking". Dans ces conditions, même comme le vendeur ne pourra ni identifier le client, ni avoir de l'information sur ses données démographiques, son comportement d'achats ou ses votes, il pourra tout de même savoir que le client a précédemment utilisé son système.

Par ailleurs, quoi qu'il existe plusieurs techniques d'hybridation pour les systèmes de recommandation [44], nous avons choisi d'illustrer notre architecture avec une technique que nous avons baptisée hybridation *napolitaine*, et qui est en fait un modèle simplifié de la technique d'hybridation mixte (Section 1.6.4). L'hybridation napolitaine consiste à rassembler les sorties des trois techniques de filtrage utilisées et à présenter le résultat au client. En d'autres termes, la recommandation "napolitaine" se résume à la mise en commun des listes d'éléments recommandés DR, CR, et FR, issues de chaque *saveur* (démographique, collaborative, par contenu). Nous présentons donc au client la recommandation "napolitaine" $DR \cup CR \cup FR$. Nous avons également introduit la notion d'essence *combinée*, formée des essences DE, CE et FE, issues de chaque technique. Comme chacune de ces essences contient des types de données différents ("saveur") et qu'elles ne sont pas mélangées, nous faisons référence à cette essence combinée sous

[12] Les données privées restent toutefois préservées.

l'appellation d'*essence napolitaine*.

Nous allons à présent décrire en détail les procédures de filtrage dans ALAMBIC. Le processus d'exécution est illustré dans la figure 5.4 ; les différentes interactions y sont matérialisées par des flèches et doivent être considérées de gauche à droite. Les différentes étapes d'exécution suivent :

Initialisation

1. Le vendeur donne le certificat de clé publique de l'agent Alambic au client, ce qui permet à ce dernier de prendre connaissance de la clé publique de l'agent Alambic et de s'assurer qu'elle est légitime[13].
2. Dédignons par π le profil démographique du client et ψ son pseudonyme. Le client crée une chaîne $\sigma = \pi || \tau$, qui contient son profil démographique et sa clé secrète, τ, que l'agent Alambic utilisera plus tard pour masquer les recommandations qui lui seront faites. Le client chiffre ce profil combiné, σ, avec la clé publique de l'agent Alambic, A, et obtient $E_A(\sigma)$. Le client envoie son pseudonyme, ψ, ainsi que $E_A(\sigma)$ au vendeur.
3. Le vendeur insère ψ et $E_A(\sigma)$ dans sa base de données de profils chiffrés. Il fait suivre $E_A(\sigma)$ à l'agent Alambic. [À partir de cet instant et pour ses futures connexions au système, le client se sert de ψ pour communiquer avec le vendeur ; c'est pourquoi nous nous sommes concentré sur le scénario de tolérance totale au tracking.]
4. L'agent Alambic a reçu $E_A(\sigma)$ de la part du vendeur. Il déchiffre $E_A(\sigma)$ en se servant de sa clé privée, A', et obtient le profil démographique du client, π, et sa clé secrète, τ.

Filtrage démographique

1. À partir de π, l'agent Alambic calcule la distance entre les données démographiques du client et le centroïde de chaque agrégat (cluster), pour trouver

[13]Via l'autorité de certification auprès de laquelle le vendeur a déposé son certificat de clé publique.

le cluster le plus proche, $C = C_j(\pi)$, pour un certain $j \in \{1, \ldots, n\}$, où n est le nombre total de clusters dans l'essence démographique (tableau 5.4). L'agent Alambic produit alors une liste, *DR* $= \{p_{d,1}, p_{d,2}, \ldots\}$ d'index anonymes que les clients appartenant à C ont choisi par la passé. (L'indice "*d*" fait allusion au mot "démographique".) Ceci constitue le volet recommandation basé sur le filtrage démographique de notre système de recommandation. Rappelons que même si cette liste *DR* est produite par l'agent Alambic, celui-ci ne possède pas la clé Σ nécessaire pour comprendre de quels produits il s'agit.

Filtrage collaboratif

1. L'agent Alambic peut extraire les votes du client de pseudonyme ψ, à partir de la base de données des votes chiffrés de tous les clients. Pour éviter que le vendeur sache quels produits ont reçu les votes d'un client donné, chaque ligne de votes est chiffrée séparément.
2. L'agent Alambic calcule alors des prédictions de votes sur les produits que le client n'a jamais évalués (à travers des votes). À partir de ces prédictions, l'agent Alambic produit une liste *CR* $= \{p_{c,1}, p_{c,2}, \ldots\}$ d'index anonymes qui correspondent aux prédictions de valeurs plus élevées. (Voir section 1.6.3.1 pour des exemples de fonctions qui calculent les corrélations et les prédictions.)

Filtrage par contenu

1. En se servant de ψ, l'agent Alambic peut obtenir, de la base de données des votes chiffrés, la liste des produits que le client a eu à évaluer par le passé. Il peut alors consulter la base de données des similarités entre produits, dans le but d'obtenir la similarité de chaque produit que le client a évalué par le passé avec chacun des produits qu'il n'a jamais évalués. Il calcule par la suite des prédictions de votes pour les produits que le client n'a jamais évalués. À partir de ces prédictions, l'agent Alambic produit une liste, *FR* $= \{p_{f,1}. p_{f,2} \ldots\}$, formée d'index anonymes qui correspondent à des prédictions de valeurs plus élevées. (Voir section 1.6.3.2

pour des exemples de fonctions qui calculent les similarités et les prédictions.)

Hybridisation

1. L'agent Alambic produit une liste, $P = CR \cup DR \cup FR = \{p_1, p_2, \cdots, p_u\}$, formée d'index anonymes et consistant en l'union de toutes les composantes de l'essence napolitaine.

2. Il choisit au hasard deux clés secrètes δ et ζ. La première sert à protéger la procédure d'indexation du vendeur contre la concurrence. La deuxième permet de protéger la liste d'index anonymes qui est envoyée au client : cette protection est nécessaire car nous avons à faire ici à la "tolérance zéro sur les données privées". Soit S un algorithme public de chiffrement symétrique (connu de toutes les parties). L'agent Alambic applique un double chiffrement, $S(S(p_i, \delta), \zeta)$, à chacun des index anonymes $p_i, i = 1, \cdots, u$ issus de P. Il obtient alors une nouvelle liste

$$S(S(P, \delta), \zeta) = \{S(S(p_1, \delta), \zeta), \ldots, S(S(p_u, \delta), \zeta)\}.$$

   L'agent Alambic utilise sa clé secrète τ qu'il a reçue du client (étape 2 de la phase d'initialisation) pour chiffrer ζ en $S(\zeta, \tau)$. Il envoie à la fois $S(\zeta, \tau)$ et $S(S(P, \delta), \zeta)$ au client. Finallement, il envoie aussi δ au vendeur.

3. Le vendeur calcule $S(p_k, \delta)$ pour chacun des index anonymes $p_k, k = 1, \ldots, t$, où t désigne le nombre total d'index anonymes du catalogue (tableau 5.5). Il exécute alors la procédure *SELECT* (figure 5.4), destinée à créer une table T contenant des valeurs $S(p_k, \delta), k = 1, \ldots, t$, servant de clés de recherche, et les index de produits et descriptions associés. En d'autres termes, la table T contient la même information que le catalogue initial, mais est indexée par des index anonymes masqués par δ.

4. Le client déchiffre $S(\zeta, \tau)$ et obtient $\zeta = S^{-1}(S(\zeta, \tau), \tau)$. Il calcule ensuite $S(P, \delta) = S^{-1}(S(S(P, \delta), \zeta), \zeta) = \{S(p_1, \delta), \ldots, S(p_u, \delta)\}$.

5. Le client et le vendeur exécutent la procédure *RETRAIT* (figure 5.4) afin que le

TAB. 5.5 – Structure du catalogue du vendeur.

Index anonyme	Index du produit	Description	...
p_1	s_1	...	
p_2	s_2	...	
...	...	...	
p_t	s_t	...	

Remarque : Les index anonymes p_i des produits ne permettent pas à eux seuls d'identifier les produit en question. Ils sont associés des index s_i publiquement connus (ex. : une suite de chiffres) à travers une fonction de correspondance Σ connue seulement du vendeur. Ces deux index, p_i et s_i, sont accompagnés de la description du produit, et d'autres caractéristiques du produit (ex. : prix, taille, etc.).

client obtienne le détail sur les produits faisant partie de $S(P, \delta)$, mais sans révéler au vendeur les index anonymes correspondants, encore moins le catalogue tout entier au client. Ceci est un cas spécial de calcul bipartie sécuritaire ou STPC (*Secure Two-Party Computation*), entre le client et le vendeur (section 2.4.9).

Il faut noter que dans le cas d'une "tolérance totale sur les données privées", il suffirait que le client envoie toute la liste $S(P, \delta)$ de ses index masqués au vendeur. Ce dernier ferait alors une simple comparaison de cette liste avec sa table T pour obtenir des produits à recommander au client. En d'autres termes, dans ce cas, la procédure *RETRAIT* consiste simplement en la sélection des produits de T qui ont un index chiffré et anonyme dans $S(P, \delta)$.

Feedback du client

1. Soit $S(P', \delta)$ l'ensemble constitué d'index anonymes des produits que le client a achetés (feedback implicite) ou a évalués ou encore pour lesquels il a indiqué sa préférence (feedback explicite). Le client envoie $S(P', \delta)$ à l'agent Alambic pour fins de mise à jour de l'essence. $S(P', \delta)$ est envoyé sous une forme chiffrée $E_A(S(P', \delta))$ (une fois de plus, à cause de la "tolérance zéro sur les données privées" du client).

Mise à jour

La mise à jour (*MAJ*, figure 5.4) de l'essence est basée sur la liste $S(P', \delta)$ que l'agent Alambic reçoit en guise de feedback du client. À partir des éléments de cette liste, les cas suivants peuvent se présenter :

1. $S(P', \delta) = S(P, \delta)$: Le client est satisfait par tous les produits et/ou a explicitement évalué les produits associés aux index anonymes chiffrés qu'il a reçus de l'agent Alambic.
2. $S(P', \delta) \subset S(P, \delta)$: Le client a évalué quelques produits et/ou est partiellement satisfait par les produits associés aux index anonymes chiffrés qu'il a reçus de l'agent Alambic.
3. $S(P', \delta) \setminus S(P, \delta) \neq \emptyset$: Les produits que le client a évalués et/ou ceux qui l'ont satisfait ne sont pas tous issus de la liste qu'il a reçue de l'agent Alambic. La liste $S(P', \delta)$ contient des produits additionnels que le client a obtenus par d'autres moyens : par exemple, à travers des conseils d'un ami l'invitant à se procurer un produit bien précis. Dans ce cas, une description détaillée du produit aura été donnée au client, incluant son index anonyme et chiffré $S(p_i, \delta)$, pour qu'il l'ajoute à sa liste $S(P', \delta)$; ce qui lui permet alors de fournir un feedback à l'agent Alambic.

Dans tous les cas qui précèdent, les modifications nécessaires se situent au niveau :

- *de l'essence démographique* : la liste des produits associés au cluster auquel le client appartient est mise à jour par l'agent Alambic. Cela provoque un changement en C lui-même (changement du centroïde, de la densité et du rayon) et pourrait même provoquer la fusion, la séparation ou la création de nouveaux clusters. En ce qui nous concerne, tout ce qui importe est que toutes ces modifications peuvent être intégrées dans le code de l'agent Alambic et ne requièrent que la liste $S(P', \delta)$ comme donnée d'entrée ;
- *des essences collaborative et par contenu* : pour chaque produit issu de la liste $S(P', \delta)$, l'agent Alambic met à jour ou insère les votes éventuellement exprimés par le client dans la base de données de votes chiffrés.

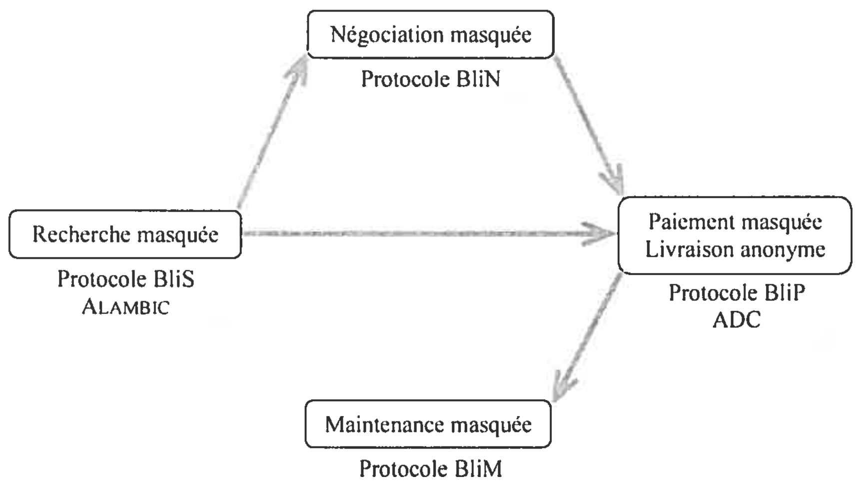


FIG. 5.5 – Vue d'ensemble du modèle BCBB.

5.9 Synthèse du modèle BCBB

Nous avons présenté notre modèle BCBB à travers la description des protocole BliS, BliN, BliP et BliM, ainsi que le système ALAMBIC et les centres de livraison anonyme (ADC). La figure 5.5 donne une vue synthétique de ce modèle.

1. Le client et le vendeur initient le processus d'achat/vente avec la recherche masquée effectuée par le client dans le catalogue du vendeur. Pour cela, les deux exécutent le protocole BliS. Toutefois, ce protocole tout seul ne permet pas au client de trouver tous les produits qui pourraient l'intéresser, surtout qu'il n'apprend qu'un prix par session. Pour résoudre ce problème, deux voies sont exploitables :

   (a) **BliS étendu** : nous pouvons légèrement modifier BliS pour en faire une version étendue, susceptible de retourner plus d'une étiquette de prix. Pour cela, il suffit de ne plus limiter le vendeur à choisir le prix le plus élevé s'il est en présence de deux lignes de produits identiques relativement à l'ensemble des caractéristiques choisies par le client (voir la description de l'ensemble Q dans la section 5.3). On se souvient alors que le vendeur se servait de $\wp$

pour masquer l'étiquette de prix. Dans la version étendue de BliS, un même $\wp$ devrait servir à masquer toutes les étiquettes de prix issues des lignes de produits identiques (relativement à Q). De plus, il faudrait permettre au client d'avoir une description plus détaillée du produit excluant uniquement les caractéristiques jugées sensibles par le vendeur ; ce qui donnerait alors une visibilité plus large au client pour décider de l'étiquette de prix à demander.

(b) **ALAMBIC** : Le système ALAMBIC apporte tous les avantages qu'ont les systèmes de recommandation à proposer au client des produits issus des calculs de corrélation, similarité et de prédiction en rapport avec ses données démographiques, ses habitudes d'achats, etc. Le plus intéressant est qu'il le fait tout en préservant la vie privée du client. En fait, ALAMBIC est un autre moyen permettant de fouiller de façon masquée dans le catalogue du vendeur. En se rappelant que l'agent Alambic envoie au client une liste d'index (anonymes), on comprend que ces derniers peuvent servir en lieu et place des index calculés dans le protocole BliS (étapes 3 pour le client et 4 pour le vendeur). En d'autres termes, on peut remplacer les étape 3 et 4 de BliS par l'exécution du système ALAMBIC et continuer l'exécution de BliS, en tenant compte, si on veut, de la notion de BliS étendu que nous venons juste de décrire.

En somme, la meilleure approche pour faire la recherche masquée devra prendre en compte une version étendue du protocole BliS et intégrer le système ALAMBIC. Après la recherche masquée, le client et le vendeur peuvent continuer avec les étapes suivantes :

2. Passer à la phase de négociation masquée en exécutant le protocole BliN,
3. Ou alors passer à la phase de paiement masqué et de livraison anonyme en exécutant BliP. Toutefois, la livraison des produits physiques est un cas particulier et la solution est l'utilisation d'un centre de livraison anonyme (ADC).
4. Dans tous les cas, la dernière phase du modèle BCBB est la maintenance (ou service après-achat) anonyme, ce qui se résume en l'exécution du protocole BliM.

5.10 Discussion sur le modèle BCBB

Dans cette section, nous décrivons quelques limites et inconvénients relatifs au modèle BCBB.

5.10.1 Protocole BliS

Le principal inconvénient avec notre approche est sa complexité en communication et en temps de calcul : le client et le vendeur doivent effectuer un nombre significatif d'opérations afin de masquer leurs questions et réponses, respectivement, et le vendeur doit envoyer beaucoup plus d'informations (chiffrées) que ce qui a vraiment de l'intérêt pour le client. En effet, si le catalogue du vendeur contient l produits, alors la complexité (pire cas) en communication et celle en temps de calcul sont chacune de l'ordre de $O(l)$ pour chaque ronde de BliS.

5.10.2 Protocole BliN

Le protocole BliN est de type bilatéral (juste deux interlocuteurs). Toutefois, il peut être modifié pour tenir dans un environnement multilatéral telles que les enchères électroniques. Cette modification fait partie de nos travaux futurs. En outre, BliN peut être utilisé dans plusieurs types de e-commere : B2B (ex. : une compagnie vend des secrets à une autre compagnie), B2C (ex. : un vendeur vend les produits aux clients—le cas de notre Alice et Sir Bob), etc.

Comme dans le cas de BliS, le principal inconvénient avec BliN est sa complexité de communication et de calcul.

5.10.3 Centres de livraison anonyme

Nous avons utilisé la notion de mix-nets de Chaum pour mettre sur pied les ADC avec plus précisément l'utilisation des messages mixtes. Dans la version originale de la technique des mix-nets, le contenu d'un courriel est à la fois inconnu et inaltérable par les intermédiaires. Ceci met en relief une certaine faiblesse de l'application de cette technique dans le monde physique où les paquets peuvent être ouverts, examinés et refermés.

En l'occurence, le vendeur pourrait faire collusion avec l'agent de livraison cible et obtenir par exemple une photo du client ou de son représentant. L'agent associerait alors la photo au contenu du paquet qu'il a eu le temps de consulter et ceci constituerait une menace à la vie privée du client. Néanmoins, cela ne pourrait avoir lieu qu'avec une faible probabilité car l'utilisation du message mixte ne révèle pas l'agent de livraison cible au vendeur. En fait, aucun agent de livraison ne pourrait prédire à l'avance qu'il a été choisi comme cible.

Une autre attaque plus subtile consisterait en ce que le vendeur fasse collusion avec un ou plusieurs agents de livraison pour que ceux-ci ouvrent les paquets au fur et à mesure qu'ils circulent dans le système de mixage. Pour contrer cette attaque, nous avons besoin d'étudier la faisabilité du point de vue technologique d'un emballage qui ne pourrait être ouvert sans possibilité d'être détecté, mais cette issue reste difficile à appliquer car aucun processus d'emballage n'est sous le contrôle du client (destinataire) ! Par conséquent, chaque agent de livraison devrait vérifier que son prédécesseur n'a pas altéré l'emballage du paquet. Les détails vont au-délà du travail effectué dans cette thèse. Nous l'avons donc classé dans la rubrique des recherches futures. Nous avons également laissé comme perspectives à venir le développement d'une procédure permettant de détecter si un agent de livraison a détourné le paquet pour ses propres besoins. Pour ce dernier point, une solution consisterait à créer un environnement dans lequel les clients exprimeraient leur confiance à travers des évaluations (votes) des agents de livraison de l'ADC.

Aussi, on ne saurait exclure la possibilité que tous les agents d'un ADC soient totalement corrompus, auquel cas les clients seraient par exemple photographiés sans cesse. Des ADC corrompus pourraient ensuite collaborer pour retracer les cas de transfert de paquets entre eux. Cette situation sort aussi du cadre de cette thèse.

Un autre problème est celui de l'utilisation des outils de tracking, comme par exemple, la technologie RFID (*Radio Frequency IDentification*) [214]. Cette technologie peut aider à surveiller les produits à distance et à espionner les individus propriétaires de ces produits. Dans le cadre des ADCs, on peut supposer que chaque agent de livraison dispose d'un ensemble de moyens "anti-étiquettes" lui permettant de détecter la présence des étiquettes de RFID et de les désactiver. Cette solution dépend alors de la technologie en ce sens que l'association "étiquettes/anti-étiquettes" est comparable à celle virus/anti-virus, spam/anti-spam, cryptographie/cryptanalyse, etc.

Par ailleurs, le processus de livraison peut nécessiter un long délai. Toutefois, cette limite peut être considérée comme étant le prix à payer pour préserver la vie privée, comme c'est d'ailleurs le cas avec plusieurs autres solutions (forage de données préservant la vie privée, transfert inconscient, etc.) qui protègent toutes ou une partie des données privées du client.

5.10.4 Système ALAMBIC

Le système *Alambic* est basé sur un tiers de semi-confiance, l'agent Alambic. La principale limite est liée au fait que le code de l'agent Alambic est chiffré et obfusqué. Jusqu'à ce jour, il n'existe pas de preuves formelles pour la sécurité des techniques d'obfuscation de codes. Toutefois, le fait de pouvoir mettre cet agent dans une plateforme autre que celle du client éliminerait cette faiblesse.

Par ailleurs, dans le cas d'une "tolérance zéro sur les données privées", la procédure *RETRAIT* se réduit à BliS ou, d'une manière générale, à une implémentation du STPC comme, par exemple, SPIR (section 3.9.3.3) ; ce qui entraînerait des coûts élevés en temps de calcul et en complexité de communication.

5.11 Conclusion

Dans ce chapitre, nous avons présenté dans le détail les protocoles du modèle BCBB que nous avons créé afin de lutter contre la violation de la vie privée en commerce électronique. Ce modèle se compose de quatre phases qui vont de (1) la recherche de produit au (4) service après-achat, en passant par (2) la négociation et (3) le paiement et la livraison. Toutes ces phases se font de manière, d'une part, à préserver la vie privée du client à tout moment et, d'autre part, à protéger les données que le vendeur juge très sensibles. Pour cela, nous avions défini au chapitre 4 une classification des données du client et du vendeur, et y avions associé une gestion de la "tolérance" (section 4.2).

En se servant de cette description et de nombreuses primitives cryptographiques (chapitre 2), nous avons pu décrire, dans le chapitre qui s'achève, tous les protocoles et systèmes nécessaires au BCBB. Pour solutionner (1), nous avons introduit le protocole BliS, ainsi que le système ALAMBIC, tandis que (2) est résolu par le protocole BliN. Quant à (3), nous avons décrit le protocole BliP et surtout les centres de livraison anonyme (ADC), utiles à la livraison des produits physiques. Pour (4), enfin, le protocole BliM peut être utilisé pour recourir aux services après-achat. Nous avons également discuté des propriétés avantageuses et des limites à utiliser le modèle BCBB.

Nous pensons que ce modèle est le plus approprié à la lutte contre la violation de la vie privée en commerce électronique, car il permet au client de contrôler ses données privées tout le long de ses échanges avec le vendeur. En outre, il permet aussi au vendeur de mieux protéger ses intérêts, grâce au contrôle qu'il fait sur ses données sensibles. Ceci cadre donc avec l'objectif que nous nous sommes fixés dans le cadre de cette thèse, à savoir : "protéger la vie privée des clients et les données sensibles des vendeurs en commerce électronique".

Nous avons laissé l'analyse formelle des différents protocoles présentés dans ce chapitre comme travaux futurs.

CONCLUSION

Nous avons traité dans cette thèse du problème de la vie privée en commerce électronique. Plus précisément, nous avons présenté notre approche pour résoudre les problèmes qui concernent la protection de la vie privée des clients et des données sensibles des vendeurs en commerce électronique. Après avoir fait un survol du commerce électronique dans son ensemble, nous avons présenté les primitives et systèmes cryptographiques dont certains nous ont permis de définir des protocoles utiles à notre approche. Nous avons par la suite décrit la problématique relative à la vie privée sur Internet en général et, en particulier, en commerce électronique. En partant de cette problématique, nous avons survolé l'état de l'art des méthodes de protection de la vie privée avant d'introduire notre approche.

Notre approche a consité en la création d'un nouveau modèle, le BCBB (*Blind Customer Buying Behaviour*), afin d'aider le client à :

1. rechercher le produit capable de satisfaire son besoin dans le catalogue du vendeur de façon *masquée*, c'est-à-dire sans que le vendeur sache explicitement ce que le client recherche, mais qu'il soit en mesure de donner des réponses plausibles au client ;

2. négocier, toujours de façon masquée, le prix de vente final ainsi que les conditions de vente du produit ;

3. payer le vendeur sans laisser de trace et recevoir de manière anonyme la livraison du produit acheté, qu'il soit numérique ou physique ;

4. bénéficier des services après vente de manière masquée.

Les quatre étapes ci-dessus ont été décrits respectivement sous forme de protocoles BliS (*Blind Search*), BliN (*Blind Negotiation*), BliP (*Blind service and delivery*) et BliM (*Blind Maintenance*).

Pour compléter le protocole BliS qui se limite à une recherche basée sur les caractéristiques du produit, nous avons introduit ALAMBIC, un système de recommandation qui préserve la vie privée des clients tout en protégeant les données sensibles du vendeur. Aussi, pour permettre la livraison anonyme des produits physiques, nous avons introduit les centres de livraison anonyme. Le cas des biens électroniques quant à lui trouve la solution dans la navigation anonyme sur Internet.

Notre modèle (BCBB) est différent du standard CBB (*Customer Buying Behaviour*) dans lequel les six étapes qui le composent ne mettent pas l'accent dans la protection de la vie privée des clients. Il protège aussi les données sensibles du vendeur.

Tous ces protocoles et systèmes demandent à être implémentés dans l'industrie. Aussi, il est important de tenir compte de toutes les faiblesses qu'ils présentent (section 5.10) afin de les améliorer. Nous pensons que ces deux éléments (implémentation et amélioration) sont étroitement liés et comptons donc y concentrer nos efforts en termes de travaux futurs.

BIBLIOGRAPHIE

[1] J. Abbate, *Inventing the Internet*, MIT Press, Cambridge, 1999

[2] N. R. Adam, O. Dogramaci, A. Gangopadhyay and Y. Yesha, *Electronic commerce : Technical, Business, and Legal Issues*, Prentice Hall, 1998.

[3] C. Adams and S. Lloyd, *Understanding PKI : Concepts, Standards, and Deployment Considerations*, 2nd Edition, Addison-Wesley, 2002.

[4] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems : A Survey of the State-of-the-Art and Possible Extensions", *IEEE Transactions on knowledge and data engineering*, **17**(6) :734–749, 2005.

[5] C.C. Aggarwal, J.L. Wolf, K-L. Wu and P.S. Yu, "Horting Hatches an Egg : A New Graph-Theoretic Approach to Collaborative Filtering", *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 201–212, 1999.

[6] R. Agrawal, T. Imielinski and A. Swami, "Mining Association Rules between Sets of Items in Large Databases", *Proceedings of ACM SIGMOD-93*, pp. 207–216, 1993.

[7] B. Aiello, Y. Ishai and O. Reingold, "Priced oblivious transfer : How to sell digital goods", *Advances in Cryptology : Proceedings of Eurocrypt 01*, pp. 119–135, Springer-Verlag, 2001.

[8] E. Aïmeur, G. Brassard and F. S. Mani Onana, "Blind sales in electronic commerce", *Proceedings of the 6th International Conference on Electronic Commerce (ICEC'04)*. Delft, The Netherlands, pp. 148–157, 2004.

[9] E. Aïmeur, G. Brassard and F. S. Mani Onana, "Blind negotiation in electronic commerce", *Proceedings of Montreal Conference on eTechnologies 2005*. Montréal, Canada, pp. 35–43, 2005.

[10] E. Aïmeur, G. Brassard and F. S. Mani Onana, "Blind electronic commerce", *Journal of Computer Security*, 2006. *À paraître*.

[11] E. Aïmeur, G. Brassard and F. S. Mani Onana, "Privacy-preserving physical delivery in electronic commerce ", In *IADIS International Conference e-Commerce 2005*, pp. 25–33, Porto, Portugal, 2005. *Best Paper Award.*

[12] E. Aïmeur, G. Brassard and F. S. Mani Onana, "Secure Anonymous Physical Delivery", *WWW/Internet Journa*, 2006. *À paraître.*

[13] E. Aïmeur, G. Brassard, J. M. Fernandez and F. S. Mani Onana, "ALAMBIC : A privacy-preserving recommender system for electronic commerce", *ACM Transactions on Information Technology. À paraître.*

[14] E. Aïmeur, G. Brassard, J. M. Fernandez and F. S. Mani Onana, "Privacy-preserving demographic filtering", In *Proceedings of the 21st Annual ACM Symposium on Applied Computing*, pp. 872–878, Dijon, France, 2006.

[15] E. Aïmeur, A. Frantz-Desmarais and F. S. Mani Onana, "User-supervised recommender systems", *International Journal of Pattern Recognition and Artificial Intelligence*. Soumis.

[16] E. Aïmeur and F. S. Mani Onana, "Better control on recommender systems", *IEEE Conference on E-Commerce Technology (CEC'06)*, pp. 297–306, San Francisco, CA, 2006.

[17] E. Aïmeur and Y. Ma, "Intelligent Agent in Electronic Commerce *XMLFinder*", *10th IEEE International Workshops on Enabling Technologies : Infrastructure for Collaborative Research , Knowledge Media Networking Workshop*, MIT, Cambridge, MA, Juin 2001.

[18] S. Allamaraju et al., "Professional Java E-commerce", Wrox Press, 2001.

[19] A. Andreasen, "Attitudes and Customer Behavior : A Decision Model", In *L. Preston (ed.), New Research in Marketing*, California Institute of Business and Economics Research, University of California, 1965.

[20] D. Asonov and J.-C. Freytag, "Almost optimal private information retrieval", Technical report HUB-IB-156, Humboldt University, Berlin, 2001.

[21] J. Bailey and Y. Bakos, "An exploratory study of the emerging role of electronic intermediaries", *International Journal of Electronic Commerce*, **1**(3) :7–20, 1997.

[22] M. Balabanovic, "An Adaptive Web Page Recommendation Service", *Proceedings of the First International Conference on Autonomous Agents (Agents'97)*, pp. 378–385, 1997.

[23] M. Balabanovic and Y. Shoham, "Fab : Content-based, Collaborative Recommendation", *Communications of the ACM*, **40**(3) :66–72, March 1997.

[24] C. Basu, H. Hirsh and W. Cohen, "Recommendation as classification : using social and content-based information in recommendation", *Proceedings of the 1998 National Conference on Artificial Intelligence (AAAI-98)*, pp. 714–720, 1998.

[25] D. Beaver, J. Feigenbaum, J. Kilian and P. Rogaway, "Locally random reductions : Improvements and applications", *Journal of Cryptology* **10**(1) :17–36, 1997.

[26] F. Bélanger and C. V. Slyke, *E-business Technologies, Supporting the Net-Enhanced Organisation*, John Wiley & Sons, 2003.

[27] A. Belleil, *e-Privacy. Le marché des données personnelles : protection de la vie privée à l'âge d'Internet*, Dunod, Paris, 2001.

[28] M. Ben-Or, S. Goldwasser and A. Wigderson, "Completeness Theorems for Non-Cryptographic Fault-Tolerant Distributed Computation", *Proceedings of the 20th STOC*, pp. 1–10, New York, 1988.

[29] M. Ben-Or and S. Rabin, "Verifiable secret sharing and multiparty protocols with honest majority", *Proceedings of the 21st STOC*, pp. 73–85, New York, 1989.

[30] M. Benyoucef, *Combined Negotiations in E-commerce : Concepts, Architecture, and Implementation*, PhD Thesis, Université de Montréal, 2002.

[31] A. Berson, S. Smith and K. Thearling, *Building data mining applications for CRM*, McGraw-Hill, 2000.

[32] J. Bettman, *An Information Processing Theory to Consumer Choice*, Addison-Wessley, 1979.

[33] E. Biham and A. Shamir, "Differential cryptanalysis of the full 16-rounds DES, Advances in Cryptology", *Proceedings of Crypto'92*, pp. 487–496, Springer-Verlag, 1992.

[34] D. Billsus and M. J. Pazzani, "User modeling for adaptive news access", *User modeling and user-adaptated interaction* **10**(2-3) :147–180, 2000.

[35] I. F. Blake and V. Kolesnikov, "Strong Conditional Oblivious Transfer and Computing on Intervals", *Proceedings of Asiacrypt 2004*, P. J. Lee (Ed.), pp. 515–529, Springer-Verlag, 2004.

[36] I. F. Blake, G. Seroussi and N. Smart, "Elliptic curves in cryptography", Cambridge University Press, 1999.

[37] M. Blum, P. Feldman and S.Micali, "Non-Interactive Zero-Knowledge and its Applications", *Proceedings of the 20th STOC*, pp. 103–112, New York, 1988.

[38] F. Boudot, B. Schoenmakers, J. Traoré, "A fair and efficient solution to the socialist millionaire's problem", *Discrete Applied Mathematics*, **111** :23–36, 2001.

[39] G. Brassard, *Cryptographie contemporaine*, Masson, 1992.

[40] G. Brassard, C. Crépeau and J.-M. Robert, "All-or-nothing disclosure of secrets", *Advances in Cryptology : Proceedings of Crypto 86*, pp. 234–238, Springer-Verlag, 1987.

[41] J. Breese, D. Heckerman and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering", *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI-98)*, pp. 43–52, 1998.

[42] C. Brodie, C-M. Karat, J. Karat and J. Feng, "Usable security and privacy : a case study of developing privacy management tools", *SOUPS '05 : Proceedings of the 2005 symposium on Usable privacy and security*, pp. 35–43, Pittsburgh, Pennsylvania, 2005.

[43] H. P. Brougham, "William Pitt, the elder, Earl of Chatham, speech in the House of Lords, *Historical Sketches of Statesmen Who Flourished in the Time of George III*, vol. 1, p. 52, 1839.

[44] R. Burke, "Hybrid Recommender Systems : Survey and Experiments", *User Modeling and User-Adapted Interaction*, **12**(4) :331–370, Hingham, MA, 2002.

[45] R. Burke, B. Mobasher and R. Bhaumik, "Limited Knowledge Shilling Attacks in Collaborative Filtering Systems", In *Working Notes of IJCAI-05 Workshop on Intelligent Techniques for Web Personalization*, pp. 17–24, Edinburgh, Scotland, 2005.

[46] H. Burkert, "Privacy-enhancing technologies : Typology, critique, vision". In *Technology and Privacy : The New Landscape*, (P. E. Agre & M. Rotenberg, éds.), MIT Press, pp. 125–142, 1997.

[47] S. Burnett and S. Paine, *RSA Security's Official Guide to Cryptography*, McGraw-Hill, 2001.

[48] C. Cachin, S. Micali and M. Stadler, "Computationally private information retrieval with polylogarithmic communication", *Advances in Cryptology : Proceedings of Eurocrypt 99*, pp. 402–414, Springer-Verlag, 1999.

[49] J. Canny. "Collaborative filtering with privacy". In *I*EEE Symposium on Security and Privacy, pp. 45–57, Oakland, CA, May 2002.

[50] J. Canny. "Collaborative filtering with privacy via factor analysis". In *P*roceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, pages 238–245, Tampere, Finland, August 2002.

[51] Y.-C. Chang, "Single database private information retrieval with logarithmic communication", `http://eprint.iacr.org/2004/036/`, dernier accès le 20 novembre 2005.

[52] C. Chassigneux, "Aterritorialité des atteintes face aux logiques territoriales de protection juridique et problème de l'absence d'homogénéité des législations protectrices (quid des safe harbor principles)", Vie privée et interconnexions : vers un changement de paradigme ? *Conférence organisée par le Programme international de coopération scientifique (CRDP / CECOJI)*, Ivry sur Seine, 5 juin 2003. `http://www.lex-electronica.org/articles/v9-2/chassigneux.pdf`, dernier accès le 20 novembre 2005.

[53] C. Chassigneux, *Vie privée et commerce électronique*, Éditions Thémis, 2005.

[54] D. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms", *Communications of the ACM* **24**(2) :84–90, February 1981.

[55] D. Chaum, "Blind Signatures for Untraceable Payments", *Proceedins of Crypto'82*, pp. 199–203, August 1982.

[56] D. Chaum, "Security without identification : Transaction systems to make Big Brother obsolete", *Communications of the ACM* **28**(10) :1030–1044, October 1985.

[57] D. Chaum, C. Crépeau and I. Damgård, "Multiparty Unconditionally Secure Protocols", *Proceedings of the 21st STOC*, pp. 11–19, New York, 1988.

[58] D. Chaum, I. Damgård and vd. Graaf, "Multiparty Computations ensuring secrecy of each party's input and correctness of the result", *Proceedings of Crypto 87*, pp. 87-119, Springer Verlag, 1987.

[59] P-A. Chirita, W. Nejdl and C. Zamfir, "Preventing shilling attacks in online recommender systems", *7th ACM International Workshop on Web Information and Data Management (WIDM'05)*, Bremen, Germany, November 5, 2005. `http://www.l3s.de/~chirita/publications/chirita05preventing.pdf`, dernier accès le 20 novembre 2005.

[60] B. Chor, O. Goldreich, E. Kushilevitz and M. Sudan, "Private information retrieval", *Proceedings of 36th Annual IEEE Symposium on Foundations of Computer Science*, pp. 41–51, 1995.

[61] R. Clarke, "Fundamentals of negotiation", `www.anu.edu.au/people/Roger.Clarke/SOS/FundasNeg.html`, 1993,
dernier accès 10 Septembre 2005.

[62] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes and M. Sartin, "Combining Content-Based and Collaborative Filters in an Online Newspaper", *Proceedings of ACM SIGIR Workshop on Recommender Systems*, Berkeley, CA, August 1999.

[63] W. J. Clinton and A. Gore, *A Framework for Global Electronic Commerce*, The White House, 1997.

www.technology.gov/digeconomy/framewrk.htm, dernier accès 10 Septembre 2005.

[64] C. Collberg, "The obfuscation and software watermarking home page", 1993. http://www.cs.arizona.edu/~collberg/Research/Obfuscation/Resources.html, dernier accès le 29 octobre 2005.

[65] C. Collberg, C. Thomborson and D. Low, "A taxonomy of obfuscating transformations", Technical Report 148, Department of Computer Science, University of Auckland, July 1997.

[66] T. Cooley, "A treatise on the constitutional limitations which rest upon the legislative power of states of the American union", 2nd edition, Callaghan & Co., Chicago, 1888.

[67] T. M. Cover. "Rates of Convergence for Nearest Neighbor Procedures". In *Proceedings of Hawaii International Conference on System Science*, IT-13, pp. 413–415, 1968.

[68] T. M. Cover and P. E. Hart. "Nearest Neighbor Pattern Classification". *IEEE Transaction on Information Theory*, IT-13, pp. 21–27, 1967.

[69] L. F. Cranor. "Privacy policies and privacy preferences". In *Security and usability*, (L. F. Cranor et S. Garfinkel, éds.), Cambridge : O'Reilly, chapitre 22, pp. 429–454, 2005.

[70] C. Crépeau, "Equivalence between two flavours of oblivious transfers", *Advances in Cryptology : Proceedings of Crypto 87*, pp. 350–354, Springer-Verlag, 1988.

[71] J. Daemen and V. Rijmen, "AES proposal : Rijndael", csrc.nist.gov/CryptoToolkit/aes/, dernier accès, 18 novembre 2005.

[72] W. Dai and R. Cohen, "Content Augmentation Aspects of Personalized Entertainment Experience", *Proceedings of the 3rd Workshop on Personalization in Future TV*, pp. 12–21, Johnstown, Pennsylvania, June 2003.

[73] A. Desmarais-Frantz and E. Aïmeur. "Community cooperation in Recommender Systems". In *P*roceedings of IEEE International Conference on e-Business Engineering (ICEBE 2005), Bejing, octobre 2005. Accepté.

[74] W. Diffie and M. E. Hellman, "New directions in cryptography", *IEEE Transactions on Information Theory* **22**(6) :644–654, October 1976.

[75] J. M. Dinant, "Les traitements invisibles sur Internet", *Cahiers du CRID (Centre de Recherches Informatique et Droit)*, Facultés Universitaires Notre-Dame de la Paix, Faculté de Droit, No. 16, pp. 271–294, Bruylant, 1999.

[76] P. Ekdahl and T. Johansson, "A New Version of the Stream Cipher SNOW", *Proceedings of the 9th Annual International Workshop on Selected Areas in Cryptography 2002*, pp. 47–61, St. John's, Newfoundland, 2002.

[77] T. ElGamal, "A public key cryptosystem and a signature scheme based on discrete logarithms", *IEEE Transactions on Information Theory*, **31** :469–472, 1985.

[78] J. Engel and R. Blackwell, *Consumer Behavior*, 4th ed., CBS College Publishing, 1982.

[79] C. Endorf, G. Schultz and J. Mellander, *Intrusion Detection and Prevention*, McGraw-Hill, 2003.

[80] Federal Information Processing Standard 186, "Digital Signature Standard (DSS)", FIPS PUB 186, U.S. Dept. of Commerce National Institute of Standrds and Technology, Gaithersburg, MD., May 1994,

[81] U. Feige, A. Fiat and A. Shamir, "Zero-Knowledge Proofs of Identity", *Journal of Cryptography*, pp. 77–95, 1988.

[82] S. Flinn and J. Lumsden, "User perceptions of privacy and security on the Web", In *Proceedings of Third Annual Conference on Privacy, Security and Trust*, pp. 15–26, St. Andrews, New Brunswick, October 2005.

[83] G. Fouchard, *E-commerce dotcorps & dotcoms - La stratégie gagnante*, Osman Eyrolles Multimédia (OEM), 2001.

[84] P-A. Fouque, "Le partage de clés cryptographiques : Théorie et Pratique", thèse de doctorat, Université de Paris 7, Octobre 2001.

[85] J. Freyne, and S. Smyth, "communities, collaboration and cooperation in personalized Web search", In *Working Notes of IJCAI-05 Workshop on Intelligent Techniques for Web Personalization*, pp. 73–80, Edinburgh, Scotland, 2005.

[86] S. Garfinkel and G. Spafford, *Web Security, Privacy & Commerce*, O'REILLY, 2nd edition, 2002.

[87] Y. Gertner, Y. Ishai, E. Kushilevitz and T. Malkin, "Protecting data privacy in private information retrieval schemes", *Proceedings of 30th Annual ACM Symposium on the Theory of Computing*, pp. 151–160, 1998.

[88] O. Goldreich, S. Micali and A. Wigderson, "How to Play any Mental Game", *Proceedings of the 19th STOC*, pp. 218–229, 1987.

[89] N. Good, J. B. Schafer, J. A. Konstan, A. Borchers, B. Sarwar, J. Herlocker and J. Riedl, "Combining Collaborative Filtering with Personal Agents for Better Recommendations", *Proceedings of AAAI-99, AAAI Press*, pp. 439–446, 1999.

[90] S. Guillem-Lessard. Générateurs de nombres aléatoires. Tutorial de cryptographie : www.uqtr.ca/ delisle/Crypto/generateurs/, dernier accès le 22 novembre 2005.

[91] R. H. Guttman, A. G. Moukas and P. Maes, "Agent-mediated electronic commerce : A survey", *Knowledge Engineering Review Journal* **13**(3) :985–1003, June 1998.

[92] C. A. Henk van Tilborg, *Encyclopedia of cryptography and security*, Springer, 2005.

[93] J. Howard and J. Sheth. *The Theory of Buyer Behavior*, Wiley and Sons, 1969.

[94] A. K. Jain, M. N. Murty and P. J. Flynn "Data clustering : a review". *ACM Computing Surveys*, **31**(3) :264–323, New York, NY, 1999. ACM Press.

[95] N. R. Jennings, P. Faratin, A. R. Lomuscio, S. Parsons, C. Sierra and M. Wooldridge, "Automated negotiation : Prospects, methods and challenges", *International Journal of Group Decision and Negotiation* **10**(2) :199–215, 2001.

[96] N. R. Jennings, S. Parsons, P. Noriega and C. Sierra, "A framework for argumentation-based negotiation", *Fourth International Workshop on Agent Theories Architectures and Languages (ATAL-97)*, Springer-Verlag, pp. 177–192, 1998.

[97] N. R. Jennings, S. Parsons, C. Sierra and P. Faratin, "Automated negotiation", *Proceedings of 5th International Conference on Practical Application of Intelligent Agents and Multi-Agent Systems (PAAM-2000)*, pp. 23–30, 2000.

[98] A. H. Karp, "Representing Utility for Automated Negotiation", Technical Report, HPL–2003–153, HP Labs, 2003.

[99] C. Kaufman, R. Perlman and M. Speciner, *Network Security : Private Communication in a Public World*, 2nd edition, Prentice Hall, 2002.

[100] L. M. Kohnfelder, *Towards a Practical Public-key Cryptosystem*, B.Sc. thesis, MIT, 1978.

[101] J. A. Konstan, J. T. Riedl and A. Jameson, "Tutorial : AI Techniques for Personalized Recommendation". In *P*roceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI-2003), Acapulco, Mexico, June 2003.

[102] B. Krulwich, "Lifestyle Finder : Intelligent User Profiling Using Large-Scale Demographic Data", *AI Magazine*, **18**(2) :37–45, 1997.

[103] M. Kumar and S. E. Feldman, "Business negotiations on the Internet", *INET98 Conference of the Internet Society*, Geneva, Switzerland, July 1998.

[104] E. Kushilevitz and R. Ostrovsky, "Replication is not needed : Single database, computationally-private information retrieval", *Proceedings of 38th Annual IEEE Symposium on Foundations of Computer Science*, pp. 364–373, 1997.

[105] E. Kushilevitz and R. Ostrovsky, "One-way trapdoor permutations are sufficient for single-server private information retrieval", Technical report CS0962, Computer Science Department, Technion, 1999.

[106] K. Lam Shyong and J. Riedl, "Shilling recommender systems for fun and profit". In *Proceedings of the 13th international conference on World Wide Web (WWW'04)*, pp. 393–402, New York, NY, August 2004.

[107] S. Laur and H. Lipmaa, "Additive Conditional Disclosure of Secrets And Applications". In *Cryptology ePrint Archive, Report 2005/378*, 2005. `http://eprint.iacr.org/2005/378`, dernier accès le 22 novembre 2005.

[108] H. Lipmaa, "Computationally private information retrieval with quasilogarithmic total communication", `http://eprint.iacr.org/2004/063/`, 2004. Dernier accès le 22 novembre 2005.

[109] I. J. Lloyd and M. Simpson, *Law on the electronic frontier*, David Hume Institute, 1998.

[110] A. R. Lomuscio, M. Wooldridge and N. R. Jennings, "A classification scheme for negotiation in electronic commerce", *Agent-Mediated Electronic Commerce : A European AgentLink Perspective*, pp. 19–33, 2001.

[111] P. Maes, "Agents that reduce work and information overload", *Communications of the ACM*, **37**(7) :31–40, July 1994.

[112] D. Martin, "Privacy Analysis for the Casual User with Bugnosis". In Security ans Usability, (L.F. Cranor et S. Garfinkel, éds.), chapitre 23, pp. 455-476, Cambridge : O'Reilly, 2005.

[113] U. M. Maurer and S. Wolf, "The relationship between breaking the Diffie-Hellman protocol and computing discrete logarithms", *SIAM Journal of Computing* **28**(5) :1689–1721, 1999.

[114] P. May, *The Business of Ecommerce*, Cambridge university Press, 2000.

[115] P. May and D. Orchard, *The Business of Ecommerce : From Corporate Strategy to Technology*, Cambridge University Press, 2000.

[116] P. Melville, R. J. Mooney and R. Nagarajan, "Content-Boosted Collaborative Filtering for Improved Recommandations", *Proceedings of Eighteenth national conference on Artificial intelligence*, pp. 187–192, Edmonton, Canada, July 2002.

[117] A. J. Menezes, *Elliptic curve public key cryptosystems*, Kluwer Academic Publishers, 1993.

[118] A. J. Menezes, P. C. van Oorschot and S. A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1996.

[119] R. C. Merkle, "Secure communications over insecure channels", *Communications of the ACM* **21** :294–299, 1978.

[120] M. Montaner, B. Lopez and J. L. De La Rosa, "A Taxonomy of Recommender Agents on the Internet", *Artificial Intelligence review*, **19**(4) :285–330, June 2003.

[121] R. J. Mooney and J. Roy, "Content-based book recommending using learning for text categorization", *Proceedings of the fifth ACM conference on Digital libraries*, pp. 195–204, San Antonio, Texas, 2000.

[122] M. Naor and B. Pinkas, "Oblivious transfer and polynomial evaluation", *Proceedings of 31st Annual ACM Symposium on the Theory of Computing*, pp. 294–303, 1997.

[123] M. Naor and B. Pinkas, "Oblivious transfer with adaptive queries", *Advances in Cryptology : Proceedings of Crypto 99*, pp. 573–590, Springer-Verlag, 1999.

[124] D. M. Nichols, "Implicit Rating and Filtering", *Proceedings of the 5th DELOS Workshop on Filtering and Collaborative Filtering*, pp. 31–36, Budapest, Hongrie, 1997.

[125] F. Nicosia, *Consumer Decision Processes : Marketing and Advertising Implications*, Prentice Hall, 1966.

[126] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, 2000.

[127] OCDE, "Organisation de Coopération et de Développement Économiques", *Lignes directrices de l'OCDE sur la protection de la vie privée et les flux transfrontières de données de caractère personnel*, Éditions OCDE, 2002.

[128] M. P. O'Mahony, N. Hurley, N. Kushmerick and G. Silvestre, "Collaborative recommendation : A robustness analysis", *ACM Transactions on Internet Technology, Special Issue on Machine Learning for the Internet*, **4**(4) :344–377, November 2004.

[129] M. J. Pazzani, "A Framework for Collaborative, Content-Based and Demographic Filtering", *Artificial Intelligence Review archive*, **13**(5/6) :393–408, 1999.

[130] D. Peppers and M. Rogers, *Le one to one : valorisez votre capital-client*, Les Éditions d'Organisation, Paris, 1998.

[131] A. Pfitzmann and M. Köhntopp, "Anonymity, Unobservability, and Pseudonymity - A Proposal for Terminology", In *Proceedings of the International Workshop on Design Issues in Anonymity and Unobservability*, pp. 1–9, Berkeley, CA, 2000.

[132] G. Plouin, J. Soyer and M-É. Trioullier, *Sécurité des architectures Web*, Dunod, 2004.

[133] H. Polat and W. Du. "SVD-based collaborative filtering with privacy". In *The 20th ACM Symposium on Applied Computing, Track on E-commerce Technologies*, pp. 13–17, Santa Fe, New Mexico, March 2005.

[134] G. Pujolle, *Les réseaux*, Édition 2005, Eyrolles, 2004.

[135] M. O. Rabin, "Digital Signatures", *Foundations of Secure Communications*, New York, Academic Press, 1978.

[136] M. O. Rabin, "Digital Signatures and public-key functions as intractable as factorization", Technical Report MIT/LCS/TR-212, Massachusetts Institute of Technology, 1979.

[137] M. O. Rabin, "How to exchange secrets by oblivious transfer", Technical Report TR-81, Harvard Aiken Computation Laboratory, 1981.

[138] M. K. Reiter and A. D. Rubin, "Crowds : anonymity for Web transactions", *ACM Transactions on Information and System Security*, **1**(1) :66–92, 1998.

[139] P. Resnick and H. R. Varian, "Recommender systems", *Communications of the ACM*, **40**(3) :56–58, 1997.

[140] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl, "Grouplens : An open architecture for collaborative filtering of netnews", *Proceedings of ACM CSCW'94 Conference on Computer-Supported Cooperative Work*, pp. 175–186, 1994.

[141] J. Revell, *Les banques et les transferts électroniques de fonds*, Paris : OCDE, 1983.

[142] V. Ring, *How to start, operate and market a freelance notary signing agent business*, Graphico Publishing, 3rd Rev edition, September 2004.

[143] R. Rivest, "MD4 message digest algorithm", RFC 1186, MIT LCS & RSA Data Security Inc, 1990.

[144] R. Rivest, "Cryptography", In *Handbook of Theoretical Computer Science*, volume A, p. 721, Elsevier, 1990.

[145] R. Rivest, "The MD5 Message-Digest Algorithm", RFC 1321, MIT LCS & RSA Data Security Inc, 1992

[146] R. Rivest, A. Shamir and L. Adleman, "A method of obtaining digital signatures and public-key cryptosystems", *Communications of the ACM*, **21**(2) :120–126, 1978.

[147] A. Rosen, *The E-Commerce Question and Answer Book : A Survival Guide for Business Managers*, 2nd edition, American Management Association, 2002.

[148] A. D. Rubin, D. Geer and M. J. Ranum, *Web security sourcebook*, John Wiley & Sons, 1997.

[149] G. Salton, *The SMART Retrieval System - Experiments in Automatic Document Processing*, Prentice-Hall, Englewood Cliffs, New Jersey, 1971.

[150] G. Salton, *Automatic Text Processing*, Addison-Wesley, 1989.

[151] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*, McGraw Hill, New York, 1983.

[152] T. Sander and C. Tschudin, "Towards mobile cryptography", *Proceedings of the IEEE Symposium on Security and Privacy*, IEEE Computer Society Press, Oakland, USA, 1998.

[153] B. M. Sarwar, G. Karypis, J. A. Konstan and J. Riedl, "Item-based collaborative filtering recommendation algorithms", In *World Wide Web*, pp. 285–295, 2001.

[154] J. Scambray, S. McClure and G. Kurtz, *Hacking exposed*, 2nd edition, McGraw-Hill, 2000.

[155] J. B. Schafer, J. A. Konstan and J. Riedl, "Recommender Systems in E-Commerce", In *EC'99 : Proceedings of the First ACM Conference on Electronic Commerce*, pp. 158–166, Denver, CO, 1999.

[156] J. B. Schafer, J. A. Konstan and J. Riedl, "E-Commerce Recommendation Applications", *Data Mining and Knowledge Discovery*, **5**(1/2) :115–153, 2001.

[157] G. P. Schneider, *Electronic Commerce*, sixth Edition, Course Technology, 2005.

[158] A. Shamir, "How to Share a Secret", *Communications of the ACM*, **22** :612–613, 1979.

[159] C. E. Shannon. A Mathematical Theory of Communication. *Bell system technical journal*, **27** :379–423, 623–656, 1948.

[160] C. E. Shannon, "Communication Theory of Secrecy Systems", *Bell system technical journal*, **28**(4) :656–715, 1949.

[161] M. Shaw, R. Blanning, T. Strader and A. Whinston, *Handbook on Electronic Commerce*, Springer, 2000.

[162] R. Shoof, "Elliptic curves over finite fields and the computation of square roots mod p", *Mathematics of Computation*, **44** :483–494, 1995.

[163] A. Scime, *Web mining : Applications and Techniques*, IGP, 2005.

[164] S. Singh, *The Code Book*, p. 120, Anchor Books, 1999.

[165] S. W. Smith and D. Safford, "Practical server privacy with secure coprocessors", *IBM Systems Journal* **40**(3) :683–695, September 2001.

[166] B. Smyth and P. Cotter, "A Personalized Television Listings Service", *Communications of the ACM*, **43**(8) :107–111, August 2000.

[167] M. Sollenborn and P. Funk, "Category-Based Filtering and User Stereotype Cases to Reduce the Latency Problem in Recommender Systems", *Proceedings of the 6th European Conference on Case Based Reasoning (ECCBR2002)*, pp. 395–405, Aberdeen, Scotland, September 2002.

[168] W. Stallings, *Business Data Communications*, 5th edition, Prentice Hall, 2004.

[169] D. R. Stinson, *Cryptography : Theory and Practice*, CRC Press, 1995.

[170] B. Suryavanshi, N. Shiri and S. Mudur, "A fuzzy hybrid collaborative filtering technique for web personalization", In *Working Notes of IJCAI-05 Workshop on Intelligent Techniques for Web Personalization*, pp. 1–8, Edinburgh, Scotland, 2005.

[171] T. Y. Tang and G. McCalla, "Smart Recommendation for an Evolving E-Learning System", *Proceedings of the 11th International Conference on Artificial Intelligence in Education (AIED'2003). Workshop on Technologies for Electronic Documents for Supporting Learning*, pp. 699–710, Sydney, Australia, July 2003.

[172] M. Teltzrow and A. Kobsa, "Impacts of user privacy preferences on personalized systems : a comparative study", *Human-Computer Interaction Series. Designing personalized user experiences in eCommerce*, pp. 315–332, Norwell, MA, 2004.

[173] H. F. Tipton and M. Krause, *Information security management handbook*, 5th edition, Auerbach, 2004.

[174] T. Tran and R. Cohen, "Hybrid Recommender Systems for Electronic Commerce", *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-00). Workshop on Knowledge-Based Electronic Markets*, pp. 78–84, Austin, Texas, July 2000.

[175] Y. Tsiounis and M. Yung, "On the security of ElGamal-based encryption", In *International workshop on Public Key Cryptography, PKC '98* (Lecture Notes in Computer Science **1431**), pp. 117–134, Yokohama, Japan, 1998.

[176] E. Turban, D King, J Lee and D. Viehland, *Electronic Commerce 2004 : A Managerial Perspective*, 3rd edition, Prentice Hall, 2004.

[177] D. Tynan, *Computer privacy annoyances*, O'Reilly, 2005.

[178] J. R. Vacca, *Public Key Infrastructure : Building Trusted Applications and Web Services*, Auerbach, 2004.

[179] VeriSign, "Building an E-commerce Trust Insfrastruture, SSL Server Certificates and Online Payment Services", Technical Brief, 2000.

[180] L. von Ahn, M. Blum, N. J. Hopper and J. Langford, "CAPTCHA : Telling humans and computers apart", *Advances in Cryptology : Proceedings of Eurocrypt 03*, pp. 294–311, Springer-Verlag, 2003.

[181] D. Watanabe, S. Furuya, H. Yoshida, K. Takaragi and B. Preneel, "A New Keystream Generator MUGI", *Proceedings of the 9th International Workshop on Fast Software Encryption 2002*, pp. 179–194, Leuven, 2002.

[182] A. Westin, *Privacy and Freedom*, Atheneum, New York, 1967.

[183] S. Wiesner, "Conjugate coding", *Sigact News*, **15** :78–88, 1983 ; original manuscript written circa 1970.

[184] A. C.-C. Yao, "Protocols for secure computation", *Proceedings of 23rd IEEE Symposium on Foundations of Computer Science*, pp. 160–164, 1982.

[185] P. R. Zimmermann, *The official PGP user's guide*, MIT Press, 1995.

Sites Web (Autres)

[186] Anonymizer.com, Online Privacy and Security, `www.anonymizer.com`, dernier accès le 30 octobre 2005.

[187] BusinessWeek Online, "Online Extra : Mobile Commerce Is Coming–Modestly, Eventually", `www.businessweek.com/magazine/content/01_20/b3732698.htm`, dernier accès le 22 novembre 2005.

[188] CattBoxx, "Internet User Statistics", `www.cattboxx.net/LatestStats.html`, dernier accès le 29 octobre 2005.

[189] Charte canadienne des droits et libertés, `laws.justice.gc.ca/en/charter/const_fr.html`, dernier accès le 29 octobre 2005.

[190] Commissariat à la vie privée du Canada, `www.privcom.gc.ca`, dernier accès le 29 Octobre 2005.

[191] Commission d'accès à l'information du Québec, `http://www.cai.gouv.qc.ca/`, dernier accès le 29 Octobre 2005.

[192] Commission Nationale de l'Informatique et des Libertés, `www.cnil.fr`, dernier accès le 29 Octobre 2005.

[193] Cookies, `wp.netscape.com/newsref/std/cookie_spec.html`, dernier accès le 24 novembre 2005.

[194] ContinentalRelay.com, *Anonymous Mail Drop & Mail Forwarding Service*, `www.continentalrelay.com`, dernier accès le 29 Octobre 2005.

[195] Cydoor, `www.cydoor.com`, dernier accès le 4 novembre 2005.

[196] Diamants infos, *The Universe of Diamond*, www.diamants-infos.com/en/, dernier accès le 22 novembre 2005.

[197] DigiStamp.com, *How a digital time stamp works*, http://www.digistamp.com/timestamp.htm, dernier accès le 22 novembre 2005.

[198] Executive Mail Drop Services, www.executivemaildropservices.com, dernier accès 5 novembre 2005.

[199] Guardster, Ltd., "Anonymous web surfing", URL : www.guardster.com, dernier accès le 22 novembre 2005.

[200] Global information infrastructure commission, www.giic.org, dernier accès le 22 novembre 2005.

[201] IETF (The Internet Engineering Task Force) , *Public-Key Infrastructure (X.509) (PKIX)* www.ietf.org/html.charters/pkix-charter.html, dernier accès le 22 novembre 2005.

[202] Ipsos Canada , www.ipsos.ca, dernier accès 5 novembre 2005.

[203] id Quantique, *Quantis - Quantum Random Number Generators* www.idquantique.com/products/quantis.htm, dernier accès le 22 novembre 2005.

[204] Kazaa, www.kazaa.com, dernier accès 4 novembre 2005.

[205] Lou Montulli, "Short bio", www.montulli.org/lou/, dernier accès 24 novembre 2005.

[206] Microsoft, *Configuring Privacy Option* www.microsoft.com/windows/ie/using/howto/privacy/config.asp, Août 2001, dernier accès 29 octobre 2005.

[207] Ministère de la justice du Canada, *Points saillants des dispositions sur la protection de la vie privée* canada.justice.gc.ca/fr/news/nr/1998/attback2.htm, dernier accès 29 octobre 2005.

[208] Network-Tools.com, Tools to trace a user, URL : www.Network-Tools.com, dernier accès le 30 octobre 2005.

[209] OCR Research Team, www.ocr-research.org.ua/, dernier accès le 22 novembre 2005.

[210] Office of the High Commissioner for Human Rights, *Déclaration universelle des droits de l'homme*, www.unhchr.ch/udhr/lang/frn.htm, dernier accès le 30 octobre 2005.

[211] Privacy and Identity Management for Europe, www.prime-project.eu.org/, dernier accès le 30 octobre 2005.

[212] Privacy International, *Big Brother Awards*, www.privacyinternational.org/bba, dernier accès le 30 octobre 2005.

[213] Proxify.com. Protect Your Online Privacy. www.proxify.com, dernier accès le 30 octobre 2005.

[214] Wikipedia. Radio Frequency Identification. en.wikipedia.org/wiki/RFID, dernier accès 19 novembre 2005.

[215] RSA LAboratories, "*RSA-155 is factored!*", www.rsasecurity.com/rsalabs/node.asp?id=2098, dernier accès le 30 octobre 2005.

[216] Sécurité publique et Protection civile Canada, "Certificat de sécurité", www.psepc-sppcc.gc.ca/prg/ns/seccert-fr.asp, dernier accès le 29 octobre 2005.

Annexe I

Vie privée au travers d'une conversation

Ceci est issu de notre travail de collaboration avec Prof. José Fernandez,

École Polytechnique de Montréal.

In the "Real World", this situation might even be comical, just imagine the following situation where a frustrated shopkeeper at a bookstore (Abbott) is dialoguing with a very particular customer (Costello). Abbott stores a large variety of books and magazines, old and new, cheap and expensive, rags and masterpieces, some tasteful, some not at all. In fact Abbott specialises in books and magazines of questionable taste, whether saucy, politically incorrect, irreverent or just plain unsavoury. In other words, stuff you wouldn't want others to know you are reading.

– Ding, ding! *Costello enters the shop*

– Well hello, How are you today?

– That's none of your business!

– Ok then... So, how can I help you?

– I want to shop, of course.

– Fine, fine. What is it that are you looking for?"

– That's none of your business! Anyways, I'll know when I find it. I'll just look through your shop..."

Costello starts rummaging through the shop. Abbott, suspicious, follows him around. Costello complains:

– Do you mind?! You are intruding on my privacy!

– Ok, ok. I'll just be at the counter if you need me.

Abbott goes back to the counter, but keeps staring at Costello, whom he definitely does not trust: he is not sure whether he is just a pickpocet or whether he might be working for the book store across the street and memorising the prices of various items...

– Will you stop staring at me!!! Really, I might have to put in a formal complaint tot he Privacy Commissioner. This is totally unacceptable!

– Listen, pal. I cannot just let you go through my store like that. Tell you what. I am just going to go the back and get myself some coffee. In the meanwhile my dog Alambic here will watch over you. Alambic, if you see this guy starting to read the magazines or write stuff down, just bark loudly, ok?

– Wouf, Wouf!!

Unfortunately, the book store is just too big and Costello gets frustrated after a while. He needs help. He calls Abbott.

– Abboooott!!!! Your store is just too big and too disorganised and I can't find what I want...

– Ok, I might be able to help you. Tell me a bit about yourself and your reading preferences, and I might be able to recommend some products you would like.

– Out of the question!!

– I though so... How about you tell me the sorts of items that you liked in the past? From that, I could either find you some other items that I have in stock that are similar to those. Also, I could look in my database of previous customers, find those who liked items similar to yours and tell you what other items they have bought that you haven't.

– You're kidding right?

– I don't see how this is going to work out...

– Hmm....

Abbott and Costello both ponder on the problem for a while. In a rare strike of genious, Costello finds an idea and suggests:

– Hey Abbott, how smart is your dog?

– Pretty smart actually. She knows the business better than I do!

– I see, it's a She-dog. And, she cannot talk or anything, can she?

– No, I don't think so...

– Perfect! I got an idea. What's the dog's name again?

– Alambic.

– Weird. Anyways, this is what we do. I whisper to Alambic what I liked and bought in the past. I also tell her what kind of guy I am...

– Yeah, I guess that's ok for you. She is smart, but I don't think she'll be able to bark away your private life anytime soon.

– Yup. And since you say she knows the business well, she can just find what's right for me, walk me to the correct aisle, and bark when I get in front of it.

– Ok, great idea. I'll just go to the back of the store and let you do your thing with the dog...

Alambic is indeed a smart dog, and helps Abbott finds what he did not even know he really wanted, a brand new copy of &@*%!* *by* !*#%$**'s (which is what every bachelor male his age is reading these days). Costello grabs the book and carefully walks back to the counter, while keeping it behind his back, away from Costello's sight.*

– Ah! Here you are. Did you find what you were looking for?

– Yes, of course. A very nice one. I'd like to pay cash for it.

– Yeah well, that's no suprise... Can I see the item please?

– Of course not!

– Aargh! I was afraid you were going to say that!! Listen pal, this is where the buck stops. You come into my store, and tell me you want to shop. You don't want to tell me who you are, what you are, what you like or what you bought. Fine!! But you have to let me know what you are taking away from my shop!!! How can I found out how much to charge you???

– That's a good point, I didn't think of that

– Wouf, wouf, wouf! Wouf, wouf! Wouf, wouf, wouf, wouf, wouf!!

– Oh shut up you! *yells Costello*

– No, wait. I think she is trying to tell us something...

– Wouf, wouf, wouf... Wouf, wouf... Wouf, wouf, wouf, wouf, wouf...

– It's three barks, two barks, five barks... Hmm... I get it!! I am suppose to charge you $3.25.

– Aaah! Good dog! Here is $5.

– Ok, Here is your change. Have a nice day now!

Costello leaves the store, already day-dreaming about his exciting find. Later on that evening, Abbott meets with another friend of his (a cryptographer) at the local watering hole and offers him his experience of the day:

– The weirdest thing happened to me today...

– Really? What?

– This weird guy comes into my shop. He says he want to shop. He does not want to tell me anything about him or about what he wants. And I am suppose to help him find what he wants.

– Yeah, I can understand his plight. Privacy is a valuable thing. Ok, how did you work it out?

– With the dog, believe it or not... He told the dog what he wouldn't tell me and she suggested something that he liked.

– That's a smart and trusty dog you have, I keep telling you.

– It gets worse. He did not want to show me what he wanted to buy. Fortunately, she managed to bark out the price.

– Wow! That's amazing. But you know what the worse part is, don't you?

– No, what's that?

– Well, this guy is so paranoid that he won't dare coming back because he will be afraid that you, or the dog for that matter, might recognize him as the same weird guy that came today. So he won't come back and all this effort will have been in vain.

– Damn!! You're right! I should have just kicked his sorry ass out of my shop the minute I saw him!

– That's throwing the baby out with the bath water. In your line of business you have to make such compromises, you should realise that. There are better solutions than kicking out pesky customers.

– Yeah? Like what?

– That's easy. You know the joke shop next door to yours?

– Yes, the one that is the in the old distillery building.

– Well, this is what you do. You enter into a joint venture with them, by which anybody wanting to go into your store can rent a face mask from them at a low fee. Because there are hundreds of different masks, the customers can pick a different one everytime they come to your shop and therefore nobody in your shop can ever tell who is who.

– Yes, but the owner of the joke shop would know...

– That's ok, the customers won't mind because he would have no idea what they are buying in your shop anyways.

– This might actually work...

– In fact if you do put in place, I even promise I will start shopping at your shop.

– Promises, promises. How would I know that you are keeping this one?

– You wouldn't, and that's the whole point...