

Analyse des comportements dans un système de transactions sur terminal mobile

Chrystel Gaber^{*†}, Romain Giot^{*}, Mohammed Achemlal ^{*†}
Baptiste Hemery[†], Marc Pasquet [†]
Pascal Urien[‡]

Résumé : Les systèmes de transactions sur mobile ont commencé à se développer récemment. L'analyse des comportements des utilisateurs de ces services permet de mieux connaître leur utilisation. Dans cette publication, nous analysons la possibilité d'obtenir des classes de comportement simples d'utilisateurs d'un système de paiement mobile à partir des traces de transactions d'un système de paiement mobile opérationnel. Une application possible de cette étude est la mise en place de moyens de détection de fraudes. Ainsi, dans des travaux futurs, nous utiliserons les résultats obtenus pour l'adaptation des algorithmes de détection de fraudes aux services de transaction sur mobile, et sur la modélisation des utilisateurs pour un simulateur de transactions et de fraudes. Nous avons ainsi analysé statistiquement le comportement de 750 000 utilisateurs ayant réalisé 4,5 millions de transactions fructueuses.

Mots Clés : Sécurité des transactions, paiement mobile, analyse de traces, modélisation de comportement, visualisation de données.

1 Introduction

Les systèmes de transactions par terminaux mobiles se démocratisent dans les pays faiblement bancarisés. Comme tout système monétaire, il est nécessaire d'effectuer une surveillance des transactions afin de détecter des transactions suspectes pouvant avoir une origine frauduleuse. Avant de mettre en place des méthodes de détection de transactions frauduleuses, il est nécessaire de comprendre la façon dont les clients utilisent le système. C'est pour cette raison que, dans cette communication, nous cherchons à déterminer les habitudes simples des clients d'un système de paiement mobile. L'acquisition de cette connaissance permettra, par la suite, de mettre en place des mécanismes de surveillance et de protection contre les transactions frauduleuses.

A notre connaissance, il n'y a pas de travaux présentant une étude comportementale d'une base de données réelle de transactions. Cette analyse présente des évidences sur le comportement général des utilisateurs, et montre également que les utilisateurs ont un comportement individuel qui leur est propre. L'analyse sécuritaire du système doit nécessairement en prendre compte.

*. France Télécom - Orange Labs, 38 rue des coutures, 14000 Caen, France {chrystel.gaber, romain.giot, mohamed.achemlal}@orange.com

†. 1 Normandie Univ, France ; 2 UNICAEN, GREYC, F-14032 Caen, France; 3 ENSICAEN, GREYC, F-14032 Caen, France; 4 CNRS, UMR 6072, F-14032 Caen, France

‡. Telecom Paristech, UMR 5141, 37/39 rue Dareau 75014, Paris, France

2 Contexte

2.1 Analyse de systèmes de paiement

Différents travaux étudient les systèmes de transaction d'argent sur mobile ainsi que leurs utilisateurs. Jack *et coll.* [JTT10] et Andrianaivo et Yartey [AY09] étudient ces types de systèmes en se basant sur différents indicateurs économiques. L'objectif de Jack *et coll.* est surtout de décrire les mécanismes économiques qui sous-tendent ces services, tandis que Andrianaivo et Yartey cherchent à étudier l'impact de ces systèmes sur les populations non bancarisées. David *et coll.* [DDA13] étudient des journaux de transaction d'un service de transaction d'argent sur mobile à l'international. Leur objectif est d'observer les flux d'argent internationaux et d'étudier les comportements économiques des utilisateurs de ce type de service. Ces travaux ont surtout pour but de confirmer ou d'infirmer des observations faites dans des études empiriques (à base de questionnaires) concernant les transferts des immigrants vers leur pays d'origine. Finalement, Ron et Shamir [RS13] étudient les comportements des utilisateurs de Bitcoin [Bit09]. Ces travaux sont ceux qui se rapprochent le plus de ce que nous souhaitons réaliser, cependant ils ne concernent pas un système de transaction d'argent sur mobile et ne ciblent pas la problématique de détection de la fraude.

Dans cet article nous ne cherchons pas à découvrir comment le service modifie l'économie ou le mode de vie des utilisateurs, nous cherchons à dégager des grandes tendances sur l'utilisation du service. Notre objectif principal est de préparer nos prochains travaux sur l'adaptation des méthodes de détection de fraudes. Un objectif secondaire est d'enrichir le simulateur de données que nous avons réalisé et proposé dans [GHA⁺13a].

2.2 Les services de transaction d'argent par mobile

Les services de transaction d'argent par mobile que nous considérons sont basés sur l'utilisation d'une monnaie électronique, également appelée m-monnaie et notée *m*. Ces services sont gérés par un opérateur de téléphonie mobile, que nous désignons dans cet article par *MNO* (pour Mobile Network Operator en anglais). Celui-ci émet la m-monnaie selon des accords qu'il a avec la banque centrale et une banque commerciale. La nature de ces accords dépend de la législation des pays où les services sont déployés.

Comme le montre la figure 1, il existe dans ce système d'autres acteurs que ceux que nous venons de mentionner. Les grossistes et les détaillants sont responsables de la distribution de la m-monnaie. Les porteurs, particuliers qui utilisent le service pour effectuer des paiements ou des transactions, et les fournisseurs de biens et de services réalisent des transactions à l'aide de la m-monnaie. Ils l'acquièrent auprès des détaillants et des grossistes en utilisant la monnaie nationale. Chacun des utilisateurs du système dispose d'un compte auprès du MNO. Les porteurs utilisent un terminal mobile pour accéder au serveur où leur compte est stocké et réaliser des paiements ou des transactions. Ils sont forcément des clients de l'opérateur de téléphonie mobile et souscrivent spécifiquement au service de paiement sur mobile. Les autres acteurs peuvent accéder à leur compte soit par un terminal mobile soit par un ordinateur.

Les services de transaction d'argent par mobile que nous considérons peuvent également être désignés par services de paiement prépayés pour mettre l'accent sur l'utilisation de la m-monnaie qui est achetée pour pouvoir être utilisée. Ces services se sont beaucoup développés ces dernières années. M-Pesa et Orange Money sont des exemples de services de transaction d'argent par mobile déjà déployés. Lancé en 2007 au Kenya, M-PESA comptait en décembre 2011, près de 19 millions de souscrivants soit environ 70% des abonnés de téléphonie mobile au Kenya [CCK12]. En juin

2012, Orange Money était déployé dans 10 pays et 14% des clients des 10 pays souscrivaient au service de transaction sur mobile [Ora12].

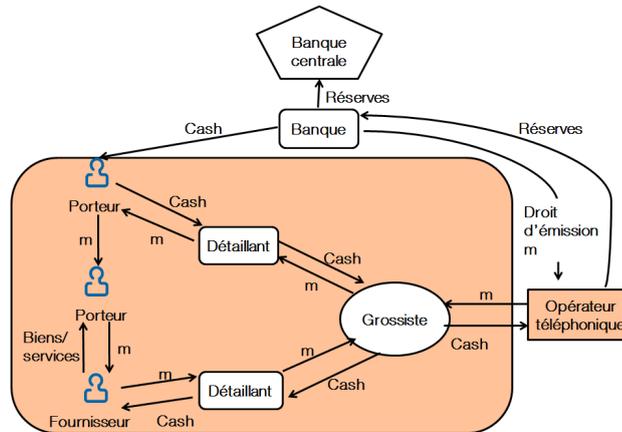


FIGURE 1: Relations entre les acteurs du paiement mobile prépayé

3 Jeu de données à notre disposition

Nous avons pu acquérir un jeu de données collecté sur 9 mois de fonctionnement d'un système de paiement opérationnel où plusieurs centaines de milliers d'utilisateurs ont effectué plusieurs millions de transactions. Ce jeu de données a été enregistré 2011, nous supposons donc qu'il ne contient pas de transaction frauduleuse.

Afin de faciliter l'analyse, nous avons effectué un nettoyage du jeu de données. 45 883 transactions enregistrées en double ont été supprimées ; cela représente 1% de la base de données que nous avons traitée, et nous n'avons pas traité les données des transactions n'ayant pas abouties (quelle qu'en soit la raison).

Les utilisateurs peuvent être classés dans les catégories suivantes : les porteurs (SUBS), les marchants (MER), les fournisseurs de services (comme l'électricité, le gaz, l'eau, ...) effectuant des factures (MERCHANT), les distributeurs de m-monnaie (DIST), les comptes opérateurs (OPT), les grossistes (WHS) et les détaillants (HOR). Les transactions sont classées dans les catégories suivantes : commission (O2C), transfert (P2P), remboursements (OPTW), enregistrement d'un nouveau client (SUBREG), retrait (CASHOUT), dépôt (CASHIN), paiement d'un marchand (MERCHPAY), paiement d'une facture (BILLPAY), réapprovisionnement d'unités téléphoniques (RC).

4 Protocole expérimental

L'objectif de cet article est d'étudier les comportements des utilisateurs, et plus particulièrement ceux des porteurs. Tout d'abord, nous analysons des statistiques générales du service (*c.-à-d.* un comportement global sur l'ensemble du système). Ensuite, nous cherchons à dégager des tendances d'utilisation du service en nous intéressant aux habitudes de paiement tant au niveau global qu'au niveau individuel.

Dans la suite de cet article, nous considérons que le comportement d'un utilisateur est constitué uniquement des transactions qui sont à son initiative, comme l'envoi d'argent à un autre particulier, le paiement d'un marchand ou le dépôt d'argent.

4.1 Statistiques générales

Les statistiques générales ont pour but de décrire et connaître le service de manière globale. Les différents indicateurs que nous avons pris en compte visent à définir l'utilisation globale du service avec le nombre total de transactions, d'utilisateurs et le montant total des transactions. Nous avons également caractérisé le service à travers l'utilisation moyenne du service pour chaque utilisateur en considérant le nombre moyen de transactions par utilisateur et le montant moyen d'une transaction.

4.2 Habitude globale d'utilisation du service

Nous avons également cherché à savoir si l'utilisation du service par l'ensemble des utilisateurs suivait des tendances. Nous avons ainsi cherché à savoir si les transactions étaient plutôt réalisées à un moment particulier de la journée et à caractériser les rôles des utilisateurs.

4.3 Habitudes individuelles d'utilisation du service

Nous considérons ici que le comportement d'un porteur est constitué d'un ensemble d'habitudes qui sont caractérisées par un type de transaction, une distribution normale bivariée sur le montant des transactions et l'écart de temps entre deux transactions successives, un début et une fin [GHA⁺13a, GHA⁺13b]. La figure 2 illustre notre vision d'une habitude. Une distribution normale bivariée est représentée ainsi que différentes transactions faisant partie de cette habitude. Afin de valider cette hypothèse, nous avons isolé les comportements de chacun des porteurs en séparant pour chacun ses transactions par types. Nous avons ainsi obtenu 634 303 listes de transactions qui représentent le comportement d'un porteur pour un type de transaction. Nous avons ensuite restreint notre étude aux listes comportant plus de 10, puis plus de 30 transactions. Nous avons ensuite appliqué un test de χ^2 [The] sur chacun de ces comportements pour déterminer si le montant et/ou la période entre deux transactions suivent une loi normale. Si le montant et la période sont normalement distribués, nous considérons que le comportement est une habitude. Il est important d'analyser individuellement le comportement des utilisateurs, car chaque utilisateur a une perception et un usage différent du système. Si les habitudes individuelles d'un utilisateur sont suffisamment stables, il pourrait être possible d'utiliser cette information pour détecter un comportement déviant pour cet utilisateur (et donc un risque de fraude).

5 Résultats expérimentaux

5.1 Statistiques générales

Le tableau 1 regroupe différents indicateurs qui décrivent de manière générale le service. La base de données que nous avons observée comprend 4 492 386 transactions, 74 933 utilisateurs et un montant total de 164 343 876 507 *m*. Le nombre moyen de transactions par utilisateur est d'environ 6 transactions pendant les 9 mois. Ce nombre est assez faible mais s'explique par le fait que la base de données que nous avons obtenue correspond à une phase de mise en place du service. Nous voyons également que les montants de transaction sont très dispersés. En effet, la variation standard du montant des transactions est très élevée et le montant médian est très différent

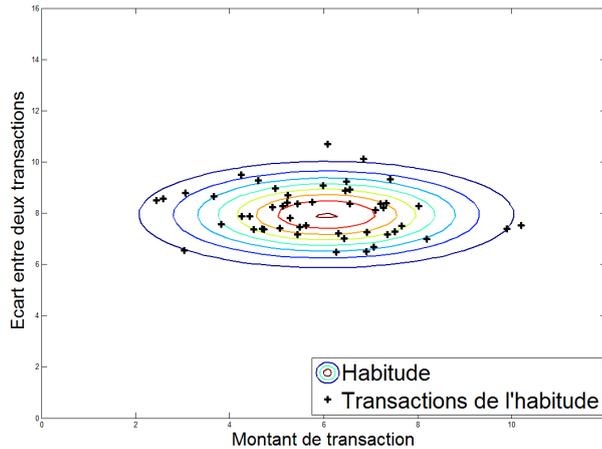


FIGURE 2: Modèle d'une habitude

du montant moyen. Il y a donc quelques transactions de montant très élevé alors que la plupart des transactions sont d'un montant inférieur à 1 000 *m*. Le montant maximal de transaction est d'ailleurs très éloigné du montant minimal.

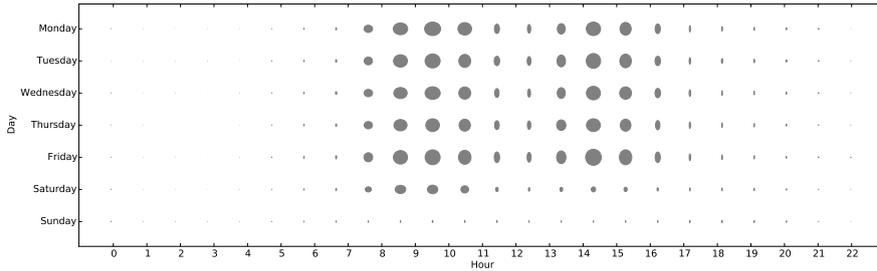
5.2 Analyse des habitudes globales d'utilisation du service

5.2.1 Analyse des interactions entre groupes d'utilisateurs

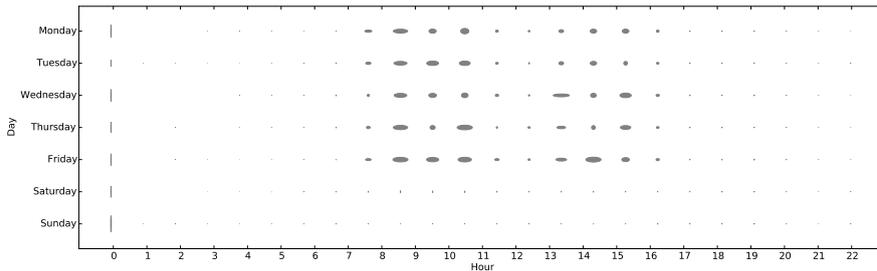
La figure 3 résume la base de données et le comportement global des différents acteurs du système. En effet, il est intéressant d'analyser le comportement de tous les types d'utilisateurs, car la fraude ne vient pas nécessairement des porteurs. Plutôt que de générer un graphe de l'ensemble des transactions à notre disposition (comme dans [RS13]), nous générons un graphe de transactions en fonction du type d'utilisateur. Chaque nœud représente un type particulier d'utilisateur, et chaque arc représente l'ensemble des transactions d'un type particulier entre chaque nœud. Le graphe a les propriétés suivantes :

TABLE 1: Statistiques générales du système de paiement étudié

| Indicateur | Valeur |
|---|--------------------|
| Nombre total de transactions | 4 492 386 |
| Nombre total d'utilisateurs | 749 933 |
| Montant total des transactions | 164 343 876 507,00 |
| Nombre de transactions par utilisateurs | 5,99 |
| Montant moyen des transactions | 36 582,76 |
| Variation standard des montants | 756 797,81 |
| Montant médian | 1 000,00 |
| Montant maximal | 300 000 000,00 |
| Montant minimal | 1,00 |



(a) Porteurs



(b) Opérateur

FIGURE 4: Résumé du comportement général des comptes clients et opérateurs en termes de montant et fréquence globale en fonction du temps

Il est donc nécessaire d'analyser individuellement les transactions de chaque utilisateur en fonction de son type et de celui du destinataire. Le nombre de transactions par client étant relativement faible, cela complexifie encore la recherche d'habitude.

5.2.2 Analyse des plages horaires d'utilisation du service

La figure 4 présente, pour les clients réguliers et les comptes possédés par l'opérateur, le nombre de transactions et le montant total de ces transactions en fonction du moment de son exécution. La largeur des ellipses est proportionnelle au nombre de transactions après une normalisation $minmax$ ⁴, et la hauteur des ellipses est proportionnelle au montant total des tractions, également après une normalisation $minmax$. Des ellipses ayant une forme la plus proche d'un cercle représentent des moments de la journée où il y a le même ratio de nombre de transactions et de montant de transactions. Des ellipses horizontales représentent des moments avec peu de transactions par rapport au comportement habituel de ce type d'utilisateurs sur les autres plages, tandis que les ellipses verticales représentent un ensemble faible de m-monnaie dépensé sur plusieurs transactions par rapport au comportement de ce type d'utilisateur.

La figure 4 montre qu'il y a deux plages majeures de transactions pour les clients : 8h00-11h00 et 13h00-16h00, qu'il n'y a quasiment aucune transaction le dimanche et moins le samedi que les autres jours de la semaine. La fréquence des transactions des habitudes est donc impactée par

4. Soustraction par la valeur minimale, puis division par la différence entre la valeur maximale et minimale

cette configuration globale. Nous retrouvons des ratios de fréquences relativement similaires dans le cas des opérateurs. Cependant, le ratio des montants est totalement différent : les transactions sont beaucoup plus nombreuses à 00h00 que dans les différentes heures de la journée. Il s'agit d'un moment où des opérations sont lancées en mode batch

Il faut noter que le comportement général des autres types d'utilisateurs (non représentés) est différent de ceux-ci : chaque catégorie d'utilisateurs a un comportement global différent.

5.3 Analyse des habitudes individuelles

Le comportement d'un utilisateur est un ensemble d'habitudes qui sont des transactions cycliques et ayant des caractéristiques similaires. Nous vérifions ici en particulier l'hypothèse sur l'aspect statistique d'une habitude, définie en 4.3, pour les porteurs uniquement. Les résultats que nous avons obtenus en considérant tous les comportements de plus de 10 transactions ou de plus de 30 transactions sont regroupés en table 2. À partir des 634 303 listes de transactions qui représentent le comportement d'un utilisateur pour un type de transaction, il en reste 23% si nous considérons les comportements sur les listes ayant plus de 10 transactions et 6% pour les habitudes de plus de 30 transactions. Nous voyons en table 2, que si l'on prend en compte tous les comportements de plus de 10 transactions, 75% des comportements suivent une distribution normale bivariée alors que pour les comportements de plus de 30 transactions, seuls 20% des comportements vérifient cette propriété. Cette différence peut avoir deux explications. Premièrement, les comportements de 10 à 30 transactions ne sont pas assez nombreux et peuvent donc être plus facilement approximés par une loi normale. La probabilité d'avoir un faux résultat positif est donc plus grande dans ce cas.

Deuxièmement, les petits comportements sont peut-être le reflet d'habitudes peu fréquentes. Si, par précaution, nous mettons ce résultat de côté, les chiffres associés aux comportements de plus de 30 transactions sont prometteurs. En effet, les 80% de comportements qui ne satisfont pas le test peuvent s'expliquer par le fait que nous avons pris une définition trop large pour les habitudes. Un utilisateur peut avoir plusieurs habitudes pour un même type de comportements. Par exemple, un utilisateur peut avoir l'habitude de faire ses courses une fois par semaine et de faire le plein d'essence une fois par mois. En observant les comportements liés aux paiements marchands, ces deux habitudes se traduiraient par la somme de deux gaussiennes, alors que le test du χ^2 ne permet d'en détecter qu'une seule. Cette analyse pourrait donc être approfondie en considérant les mélanges de gaussiennes ou en prenant en compte la notion de communauté d'intérêts pour isoler les différents comportements.

Dans tous les cas, même si il n'est pas possible de modéliser le comportement de tous les utilisateurs de cette façon : (a) nous pouvons supposer que des modélisations plus complexes seraient pertinentes, (b) un niveau de sécurité supplémentaire peut être mis en place pour ces utilisateurs.

6 Discussion

Au vu de ces résultats, on peut voir que différentes règles pourraient être mises en place afin de détecter, de façon automatique, les transactions suspectes et permettre ensuite une investigation manuelle sur l'historique des utilisateurs concernés. Les règles de détection pourraient être liées aux points suivants :

- vérification qu'un utilisateur communique avec un autre utilisateur sur un certain type de transaction tel que l'arc représentant cette transaction soit présent dans le graphe de la fi-

| | Résultats du test du χ^2 pour des comportements de plus de 10 transactions | Résultats du test du χ^2 pour des comportements de plus de 30 transactions |
|---|---|---|
| Montant et période suivent une distribution normale | 75% | 20% |
| Montant ou période suivent une distribution normale | 17% | 49% |
| Ni montant ni période ne suivent une distribution normale | 8% | 31% |

TABLE 2: Tests sur la représentation statistique d'une habitude

- gure 3, et générer une alerte si l'arc est absent ;
- la comparaison du montant de la transaction aux montants habituels de l'arc concerné de la figure 3. Une alerte est générée si la déviation du montant est trop importante (par exemple, pour les types de transactions où l'écart type n'est pas trop important, en vérifiant que la comparaison suivante est vraie : $\mu - \alpha * \sigma \leq M \leq \mu + \alpha * \sigma$, avec M le montant, μ le montant moyen de ce type de transactions, σ l'écart type de ce type de transactions et α un coefficient à paramétrer) ;
 - l'analyse des quantités et montants de transactions au cours du temps, indique que les comportements dépendent grandement du moment de la journée. Il est donc possible de mettre au point des règles générales qui génèrent des alertes lorsqu'un nombre de transactions, ou un montant total des transactions, ne correspond pas à ce qui est attendu sur la plage horaire des individus concernés ;
 - le comportement des clients peut suivre des formes spécifiques. Nous avons par exemple mis en évidence que certains comportements suivent une loi normale. La détection d'anomalie dans ce contexte est possible. Pour cela, nous devons définir un modèle de comportement normal (par exemple une loi normale) et le paramétrer à partir du comportement passé de l'utilisateur. Ce modèle pourra ensuite être utilisé pour évaluer si de nouvelles données sont proches du modèle normal ou pas. Si les nouvelles transactions sont considérées comme anormales, elles seront analysées de manière plus précise.

Bien que certaines de ces règles semblent triviales, l'analyse préalable des données permet de mieux choisir les règles métiers à appliquer et de les configurer.

En plus de ces règles simples définies manuellement, il pourrait être intéressant d'utiliser des méthodes de classification supervisée (Séparateurs à Vaste Marge [Vap99], ensemble de forêts aléatoires [Bre01]. . .) pour étiqueter les transactions. L'inconvénient des méthodes de classification est qu'elles agissent comme des boîtes noires, contrairement aux règles qui sont plus facilement compréhensibles et interprétables par un opérateur humain et qui aident à comprendre la raison des alertes.

7 Conclusion

Nous avons analysé une base réelle de transactions de paiement mobile qui contient 750 000 utilisateurs et 4,5 millions de transactions. Comme le nombre moyen de transactions est assez faible, certains résultats concernant la modélisation du comportement des utilisateurs sont difficilement interprétables. Nous avons cependant pu montrer que certains comportements suivent bien une loi normale et nous souhaitons approfondir cette étude en considérant des mélanges de gaus-

siennes. Nous avons montré que les différentes catégories d'utilisateurs ont des comportements différents en terme de catégories de contacts, types de transactions, quantité de transactions, montant de transactions, moment des transactions. Il est donc possible de mettre en place des règles génériques vérifiant le respect des contraintes identifiées.

Cependant, concernant une analyse globale des flux, à l'instar de [RS13], il serait nécessaire d'analyser le graphe complet des transactions au fil de son évolution afin de détecter des motifs complexes de fraudes. Concernant, la modélisation individuelle, il est nécessaire d'utiliser des méthodes d'analyse plus complexes afin de modéliser des comportements plus complexes (la difficulté étant la quantité de données par utilisateur potentiellement trop faible). Nos travaux futurs consisteront à résoudre ces problèmes.

Références

- [AY09] Mihasonirina ANDRIANAIVO et Charles Amo YARTEY : *Understanding the growth of African financial markets*, volume 9. International Monetary Fund, 2009.
- [Bit09] BITCOIN PROJECT : Bitcoin - an open source p2p digital currency. <http://bitcoin.org/>, 2009. Dernière visite le 25/04/2013.
- [Bre01] Leo BREIMAN : Random forests. *Machine learning*, 45(1):5–32, 2001.
- [CCK12] CCK : Quarterly sector statistics report. Rapport technique, Communications Commission of Kenya, 2012.
- [DDA13] Bounie DAVID, Diminescu DANA et François ABEL : On the effect of mobile phone on migrant remittances : A closer look at international transfers. *Electronic Commerce Research and Applications*, (0):–, 2013.
- [GHA⁺13a] Chrystel GABER, Baptiste HEMERY, Mohammed ACHEMLAL, Marc PASQUET et Pascal URIEN : Synthetic logs generator for fraud detection in mobile transfer services. In *International Conference on Collaboration Technologies and Systems (CTS 2013)*, 2013.
- [GHA⁺13b] Chrystel GABER, Baptiste HEMERY, Mohammed ACHEMLAL, Marc PASQUET et Pascal URIEN : Synthetic logs generator for fraud detection in mobile transfer services. In *Financial Cryptography*, 2013.
- [JTT10] William JACK, Suri TAVNEET et Robert TOWNSEND : Monetary theory and electronic money : Reflections on the kenyan experience. *Economic Quarterly*, 96-1(96):83–122, 2010.
- [Ora12] ORANGE : Orange money dépasse les 4 millions de clients et lance ses services en Jordanie et à Maurice. <http://www.orange.com/fr/presse/communiqués/communiqués-2012/Orange-Money-depasse-les-4-millions-de-clients-et-lance-ses-services-en-Jordanie-et-a-l-Ile-Maurice>, juin 2012. Dernière visite le 12/04/2013.
- [RS13] Dorit RON et Adi SHAMIR : Quantitative analysis of the full bitcoin transaction graph. In *Financial Cryptography and Data Security 2013*, 2013.
- [The] THE MATHWORKS, INC. : Chi-square goodness-of-fit test. <http://www.mathworks.fr/fr/help/stats/chi2gof.html>.
- [Vap99] Vladimir VAPNIK : *The nature of statistical learning theory*. springer, 1999.