Projet ANR blanc BOOLE (09 BLAN 0011)
Rapport intermédiaire à 18 mois

Avril 2011

Table des matières

1 Description des travaux effectués

1.1 F1 : Circuits et formes normales booléennes

Les activités dans ce domaine concernent d'abord la notion de dimension d'une fonction booléenne, sur laquelle a travaillé J. Vuillemin, en lien avec les formes normales d'une fonction et les diagrammes de décision.

Un deuxième travail, non encore publié, concerne la forme de certains circuits aléatoires étudiés par N. Broutin (avec O. Fawzi, McGill Univ., Canada), qui a identifié précisément la longueur des chemins, typiques et atypiques, reliant les entrées à la sortie du circuit.

1.2 F2 : Fonctions booléennes et cryptographie

Les travaux dans ce domaine ont concerné deux axes de recherche : les fonctions courbes et leurs variantes, et la construction de fonctions booléennes par la méthode des classes.

Fonctions courbes et hyper-courbes. C. Carlet et S. Mesnager ont poursuivi l'étude des fonctions courbes et hyper-courbes, dans plusieurs directions : l'obtention de nouvelles familles infinies de fonctions courbes et hyper-courbes ; le lien entre le caractère "courbe", "hypercourbe" et "semi-courbe" de plusieurs familles de fonctions booléennes ; les constructions secondaires de fonctions courbes vectorielles ; la caractérisation de fonctions "semi-courbes" en dimension paire et l'obtention de nombreuses classes infinies de fonctions semi-courbes ; l'étude d'une nouvelle classe de fonctions courbes appelée "classe H". Ils ont également établi une relation entre cette classe de fonctions courbes et une famille particulière de polynômes de la géométrie discrète ("o-polynômes"), ce qui a permis d'exploiter cinquante ans de recherche dans ce domaine particulièrement difficile et d'en déduire un grand nombre de nouvelles familles de fonctions courbes.

Méthode des classes pour les fonctions booléennes. Cette méthodologie, définie par J.-M. Le Bars et A. Viola (Montevideo, Uriguay), a été mise au point pour étudier des critères cryptographiques sur les fonctions booléennes en les décomposant récursivement en classes, et d'abord utilisée pour les fonctions 1-résilientes. Elle est appliquée ici dans deux directions :
- le codage énumératif et la génération aléatoire (J.-M. Le Bars avec A. Viola et N. Carrasco, Montevideo, Uruguay). La méthode des classes permet d'utiliser la méthode récursive, définie par Flajolet, Zimmermann et Van Cutsem en 1994 pour effectuer le codage énumératif et donc la génération aléatoire uniforme. Le véritable challenge a été de proposer des algorithmes efficaces à la fois en temps et en espace (explosion combinatoire du nombre de classes), afin de générer aléatoirement une fonction parmi les 10^{68} fonctions 1-résilientes à 8 variables.
- l'étude d'autres critères cryptographiques (J. Clément, A. Génitrini et J.-M. Le Bars, avec A. Viola). Les études concernent pour l'instant l'immunité algébrique : le plus petit degré algébrique d'un annulateur d'une fonction ou de son complémentaire. Elles ont d'abord porté sur les espaces vectoriels des annulateurs de degré algébrique fixé d'une fonction booléenne, et permis de déterminer expérimentalement toutes les dimensions possibles de ces espaces vectoriels jusqu'à 4 variables. Le travail actuel porte sur les supports de ces espaces vectoriels, dont l'étude semble une piste prometteuse pour appliquer la méthode des classes.

1.3 F3 : Satisfiabilité

Ce thème a fait l'objet de travaux des quatre partenaires. Dans le cadre des problèmes de satisfaction de contraintes (CSP), les membres du projet étudient les transitions de phase associées au comportement de formules booléennes. Ces études sont menées tant par l'analyse fine de problèmes spécifiques bien structurés que par une approche globale des CSP booléens. Dans ces deux voies de recherche, le but principal est d'identifier les structures ou les invariants qui pilotent les transitions.

Seuil de satisfaisabilité. N. Creignou, H. Daudé et R. Rossignol ont étudié, en collaboration avec U. Egly (TU Wien, Autriche), l'existence d'un phénomène de seuil pour la satisfaisabilité des formules booléennes quantifiées du type $\forall X \exists Y \varphi(X, Y)$, où φ est une conjonction de clauses contenant chacune une variable de X et deux variables de Y. Ils ont montré que la transition Sat/Insat est contrôlée par le rapport nombre de clauses/nombre de variables existentielles, et donné la localisation exacte du ratio critique. Cette étude théorique a permis de mettre au jour les structures cycliques qui pilotent la transition brusque, de satisfaisable à insatisfaisable, lorsque le nombre de variables augmente, et a été complétée par une étude expérimentale de grande envergure, rendue possible par la complexité algorithmique polynomiale du problème étudié.

Générateur de formules aléatoires. N. Creignou a entamé fin 2010 une collaboration avec U. Egly (TU Wien) et M. Seidl (Johannes Kepler University, Linz, Autriche), visant à construire un générateur

de formules aléatoires (quantifiées ou non) non nécessairement en forme normale ; ce générateur créerait des instances aléatoires en interprétant la spécification des formules souhaitées en XML. Ainsi grâce à un seul générateur on pourrait obtenir des instances aléatoires de types variés, qui pourraient en particulier être utiles pour l'évaluation des solveurs dédiés à SAT et QSAT.

Sensibilité des formules. Une fonction booléenne possède en tout point une sensibilité. N. Creignou et H. Daudé ont adapté cette notion aux formules booléennes associées aux CSP. Avec un cadre suffisamment symétrique qui permet de développer un formalisme précieux, il est possible d'énumérer (assez) simplement les formules de sensibilité donnée en une affectation. Ce résultat illustre la possibilité d'obtenir des résultats sur le comportement global des CSP partir de paramètres locaux attachés à ces problèmes.

Problèmes MaxCut et MaxXorSat. En collaboration avec C. Martinez (UPC Barcelone, Espagne), H. Daudé, V. Rasendrahasina et V. Ravelomanana étudient le nombre maximal d'arêtes d'une coupe dans un graphe aléatoire (MaxCut). En utilisant des techniques de combinatoire analytique et en développant des travaux initiés par V. Ravelomanana, ils ont obtenu des résultats en distribution qui mettent en lumière le rôle du noyau des graphes bipartis.

Pour MaxXorSat, le modèle étudié consiste à générer de manière uniforme une formule aléatoire $\mathcal{F}$ contenant M clauses, chacune de la forme $x_i \oplus x_j = \varepsilon$ avec $\epsilon \in \{0,1\}$ et $(i,j) \in [1,n]^2$. Quand $M > \frac{n}{2}$, la probabilité que la formule ainsi générée soit satisfiable tend exponentiellement vers 0. Soit $X(n,M)$ le nombre maximum de clauses satisfiables en fonction de n et de M dans une telle formule. En utilisant des méthodes issues de la combinatoire énumérative et analytique ainsi que l'étude des structures combinatoires sous-jacentes aux formules, il est possible de quantifier la variable aléatoire $X(n,M)$ de manière très précise autour de la transition de phase $M \sim \frac{n}{2}$.

Les méthodes utilisées sont amenées à s'étendre à d'autres problèmes du type Max-CSP avec des clauses portant sur 2 variables, ou encore sur les versions quantifiées telles que Max-Q-XorSat.

Propriétés monotones de l'hypercube. H. Daudé et R. Rossignol examinent le rôle des symétries sur la taille de la fenêtre de transition associée à une propriété monotone de l'hypercube. Dans le cadre plus restreint des CSP, le but est d'obtenir une amélioration du critère de transition brusque établi par Friedgut et Bourgain. Les problèmes comportant des clauses mixtes, i.e. dont le nombre de variables est borné au lieu d'être constant, peuvent être pris en compte par un produit de mesures de Bernoulli dont les paramètres peuvent varier. Ce résultat est un premier pas encourageant dans une voie de recherche en cours d'exploration.

Performances d'un solveur de contraintes. D. Gardy et C. Truchet ont mené au printemps 2010 une réflexion sur la possibilité d'évaluer les performances d'un solveur de contraintes, ce qui a conduit à définir avec X. Lorca (Ecole des Mines de Nantes) un sujet de thèse, sur lequel travaille J. du Boisberranger depuis Septembre 2010. Il s'agit d'évaluer la probabilité que l'appel à un algorithme de filtrage conduise effectivement à une réduction des domaines des variables ; le but est de fournir des outils permettant d'éviter des opérations inutiles et d'optimiser le choix des méthodes de résolution. Le premier travail a porté sur le comportement de la contrainte *AllDifferent ;* la suite devrait porter sur la prise en compte de nouvelles contraintes.

1.4 F4 : Logique quantitative

Les travaux dans ce domaine ont concerné différentes extensions de résultats antérieurs, et surtout la définition de nouveaux modèles.

Modèles de l'implication et des arbres Et/Ou. H. Fournier, D. Gardy et A. Génitrini ont finalisé l'étude quantitative des distributions sur les fonctions booléennes engendrées par les formules construites sur le seul connecteur $\rightarrow$, en collaboration avec B. Gittenberger (T.U. Wien, Autriche) et M. Zaionc (Jagiellonian Institute, Cracovie, Pologne). A. Génitrini a par ailleurs montré, avec B. Gittenberger, que l'effet Shannon (i.e., la complexité moyenne d'une fonction est aussi la complexité maximale) ne s'applique pas pour les distributions sur les fonctions booléennes engendrées par une distribution uniforme sur les arbres Et/Ou.

Modèle des arbres croissants. C. Mailler a repris le modèle des expressions booléennes, dans le cas où l'arbre sous-jacent est tiré selon la distribution de croissance aux feuilles, modèle couramment dit "des arbres binaires de recherche". Ce modèle, a priori intéressant puisque les arbres sont relativement équilibrés (hauteur en $\log n$), conduit de manière surprenante à des distributions dégénérées sur l'ensemble des fonctions booléennes.

Prise en compte des propriétés naturelles des opérateurs logiques. C. Mailler et A. Génitrini mènent actuellement, en collaboration avec B. Gittenberger et V. Kraus (T.U. Wien), une étude sur les arbres d'expressions lorsque les opérateurs $\wedge$ et $\vee$ sont associatifs et commutatifs. Ils ont déjà montré que les propriétés de commutativité ou d'associativité conduisent effectivement à des lois limites distinctes sur l'ensemble des fonctions booléennes, bien que le lien entre la probabilité $P(f)$ et la complexité $L(f)$ d'une fonction booléenne f soit conservé : $P(f) = \Theta(n^{-L(f)-1})$.

1.5 M1 : Méthodologie : méthodes de combinatoire analytique

Outre la publication d'un traité rassemblant les bases fondamentales du domaine, les travaux sur ce sujet ont porté sur plusieurs classes d'objets combinatoires : les urnes de Pólya, les arbres non planaires et les arbres enrichis.

Publication d'un livre de synthèse. Concernant les avancées en combinatoire analytique, le point le plus marquant est sans conteste la publication par P. Flajolet avec R. Sedgewick (Princeton), d'*Analytic Combinatorics,* un ouvrage majeur pour les aspects méthodologiques du projet Boole. Cette monographie regroupe pour la première fois les résultats des dernières quinze années de recherches sur les approches analytiques en combinatoire. En particulier, les propriétés asymptotiques des grandes structures aléatoires y sont traitées de façon systématique.

Urnes de Pólya. B. Morcrette a travaillé sur une classe d'urnes de Pólya à 2 couleurs équilibrées, additives et dont la fonction génératrice est algébrique. Par des méthodes de combinatoire analytique, et notamment une méthode de col faisant intervenir des cols "coalescents", on obtient des résultats sur la distribution asymptotique de l'urne, avec loi normale ainsi que loi locale limite, et grandes déviations, propriétés jusqu'alors inaccessibles par les méthodes classiques probabilistes. Ce travail poursuit l'étude des urnes de Pólya du point de vue de la combinatoire analytique, amorcée en 2006 par P. Flajolet, P. Dumas et V. Puyhaubert.

Arbres non planaires. P. Flajolet a étudié deux paramètres importants des arbres aléatoires non planaires, arbres qui interviennent dans la construction de mesures de probabilités sur les fonctions booléennes possédant des propriétés intéressantes pour la modélisation. Avec N. Broutin, il a étudié la distribution des distances dans ces arbres (les distances feuilles-racines sont cruciales pour la complexité des fonctions booléennes qu'on peut associer aux arbres ; l'aspect non-ordonné des noeuds reflète la symétrie des portes logiques par rapport aux entrées). Avec M. Bóna (Univ. de Floride, U.S.A.), il a étudié la distribution du nombre de points de symétries dans les arbres non-planaires. Ceci quantifie

précisément le biais (exponentiel) par rapport à la mesure uniforme sur les arbres planaires, pour laquelle les propriétés des fonctions booléennes engendrées ont déjà été étudiées.

Arbres enrichis. D. Gardy a étudié, avec O. Bodini (Paris 6) et B. Gittenberger (T.U. Wien), un modèle d'arbres enrichis, qui sont des classes de graphes orientés obtenus à partir d'arbres dans lesquels on ajoute des arcs de certains noeuds sources vers des feuilles. De tels objets permettent de représenter des données de manière plus souple qu'en restant dans le cadre des structures arborescentes : ainsi, pour rester dans le cadre des expressions logiques, nous pouvons prendre en compte des formules du calcul des prédicats, alors que les structures arborescentes ne permettent pas de sortir du calcul des propositions. Une première étude a permis de traiter le cas où le nombre de noeuds sources est borné sur tout chemin allant de la racine vers une feuille.

1.6 M2 : Méthodologie : méthodes probabilistes

Les travaux dans ce domaine concernent un nouveau modèle de sources dynamiques, et un opérateur sur des arbres réguliers.

Chaînes de Markov à longueur variable. B. Chauvin et N. Pouyanne étudient un modèle qui généralise les chaînes de Markov : les VLMC (Variable Length Markov Chains), chaînes de Markov à mémoire variable : en théorie de l'information, une chaîne infinie de lettres peut être vue comme une chaîne produite par une source ; on peut aussi l'étudier comme une suite infinie stochastique. Leur travail fait le pont entre ces deux points de vue, tout d'abord en établissant un cadre probabiliste pour les arbres de contexte qui servent à définir les VLMC, puis en prouvant qu'une VLMC est une source dynamique pour laquelle est donnée une transformation explicite. A terme, ce travail permettrait, par exemple, d'envisager la génération aléatoire d'une fonction booléenne dont la complexité de la représentation polynomiale est donnée : une fonction booléenne à k variables peut être représentée soit par sa table de vérité (un 2^k-uplet), soit par une fonction polynomiale à k variables (et le 2^k-uplet de ses coefficients) – dans les deux cas, par un 2^k-uplet de 0 et de 1, c'est dire un mot sur l'alphabet $\{0, 1\}$ de grande longueur. Or, grâce à une source VLMC, nous pouvons adapter les paramètres de la source afin de produire des mots dont la loi de probabilité est imposée à l'avance. Si nous parvenons faire le lien entre la complexité (classique) d'une fonction booléenne et sa complexité polynomiale, nous pouvons espérer obtenir une génération aléatoire de mots (i.e. de fonctions booléennes) dont nous connaîtrons la loi de complexité.

Opérateur sur un arbre régulier. Un travail de V. Bapst et G. Semerjian a tiré parti à la fois des méthodes probabilistes et combinatoires. Ils ont étudié un problème d'opérateurs aléatoires, plus précisément l'opérateur d'adjacence d'un arbre régulier avec poids aléatoires. Il existe une singularité essentielle dans sa densité d'états (dite queue de Lifshitz) qu'ils ont pu caractériser de manière combinatoire, à l'aide d'énumérations de marches de différents types sur l'arbre.

2 Séminaires et groupes de travail

L'activité des membres du projet a conduit à mettre en place plusieurs séminaires :

- Depuis Janvier 2011, N. Broutin organise un groupe de travail à l'Inria sur les grandes structures aléatoires. Les sujets abordés sont intimement liés à plusieurs aspects du projet Boole : limites faibles locales pour l'étude des problèmes de satisfaction de contraintes autour de la zone de transition ; aspects méthodologiques probabilistes et analytiques.
- À Marseille, dans le cadre du séminaire de l'équipe Structures Discrètes Aléatoires du LATP (UMR 6632), Hervé Daudé a organisé en Mars 2011 une demi-journée d'exposés sur le thème de

l'aléa discret avec comme invités : R. Tichy et P. Grabner (T.U. Graz, Autriche) et T. Hugel (Paris 7).

– Toujours à Marseille, N. Creignou a déposé en Novembre 2010 une demande auprès du Schloss Dagstuhl - Leibniz-Zentrum fr Informatik- pour pouvoir organiser un séminaire sur le sujet "SAT and interactions" en collaboration avec N. Galesi (Rome, Italie), O. Kullman (Swansea, U.K.) et H. Vollmer (Hanovre, Allemagne). Cette demande a été retenue par le comité scientifique en Mars 2011 ; le séminaire se déroulera fin Novembre 2012.

3 Publications

Nous indiquons ci-dessous les articles parus ou acceptés, non les articles soumis et non encore acceptés, ou en cours de rédaction.

3.1 Revues internationales à comité de lecture

1. M. Bóna, P. Flajolet : Isomorphism and symmetries in random phylogenetic trees, J. Appl. Probab., 46 (4), pages 1005–1019, 2009.
2. N. Broutin, P. Flajolet. The distribution of height and diameter in random non-plane binary trees. Random Structures and Algorithms. A paraître.
3. N. Broutin, C. Holmgren. Total path length of split trees. The Annals of Applied Probability. A paraître.
4. C. Carlet, S. Mesnager. On the construction of bent vectorial functions. Journal of Information and Coding Theory : Algebraic and Combinatorial Coding Theory. Special issue in Honour of the Retirement of Vera Pless. 1 (2), pages 133-148, 2010.
5. B. Chauvin, P. Cénac, F. Paccaut, N. Pouyanne : Context trees, variable length Markov chains and dynamical systems. Séminaire de Probabilités, Lecture Notes in Mathematics, Springer Verlag. A paraître.
6. H. Daudé, V. Ravelomanana : Random 2-XORSAT phase transition. Algorithmica, 59 (3), pages 48-65, 2011.
7. H. Fournier, D. Gardy, A. Génitrini, B. Gittenberger : The fraction of large random trees representing a given boolean function in implicational logic. Random Structures and Algorithms. A paraître.
8. H. Fournier, D. Gardy, A. Génitrini, M. Zaionc. Tautologies over implication with negative literals. Mathematical Logic Quarterly, 56 (4), pp 388-396, 2010.
9. J.-M. Le Bars, A. Viola : Equivalence classes of Boolean functions for first-order correlation. IEEE Transactions on Information Theory, pp.1247-1261, 2010.
10. S. Mesnager. A New Class of Bent and Hyper-Bent Boolean Functions in Polynomial Forms. Journal Designs, Codes and Cryptography, 59 (1-3), 265-279, 2011.
11. S. Mesnager. Bent and Hyper-bent Functions in polynomial form and Their Link With Some Exponential Sums and Dickson Polynomials. IEEE Transactions on Information Theory-IT. A paraître.

3.2 Ouvrages ou chapitres de livres

1. Flajolet, Philippe and Sedgewick, Robert, Analytic combinatorics, Décembre 2009, Cambridge University Press, pages = xiv+810.

3.3 Conférences internationales avec comité de sélection et actes

1. O. Bodini, D. Gardy, B. Gittenberger. Lambda-terms of Bounded Unary Height. International Workshop ANALCO, San Francisco (USA), Janvier 2011.
2. B. Chauvin, D. Gardy, C. Mailler. The Growing Trees Distribution on Boolean Functions. International Workshop ANALCO, San Francisco (USA), Janvier 2011.
3. G. Cohen, J.P. Flori, H. Randriambololona, S. Mesnager. On a conjecture about binary strings distribution. Proceedings of 6-th International Conference SEquences and Their Applications, SETA 2010, LNCS 6338, pp. 346-358. Springer, Heidelberg (2010).
4. A. Génitrini, B. Gittenberger : No Shannon effect on probability distributions on Boolean functions induced by random expressions. In proc. 21st International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA'10), DMTCS, Vienne, Juin 2010.
5. S. Mesnager. Hyper-bent Boolean Functions with Multiple Trace Terms. Proceedings of International Workshop on the Arithmetic of Finite Fields, M.A. Hasan and T. Helleseth (Eds.) : WAIFI 2010, LNCS 6087, pp. 97-113. Springer, Heidelberg (2010).
6. S. Mesnager. Recent Results on Bent and Hyper-bent Functions and Their Link With Some Exponential Sums. IEEE Information Theory Workshop, ITW 2010.
7. S. Mesnager, G. Cohen. On the link of some semi-bent functions with Kloosterman sums. Proceedings of International Workshop on Coding and Cryptology (IWCC 2011), pages 263-272, LNCS 6639, Springer, 2011.
8. V. Rasendrahasina and V. Ravelomanana : Limit Theorems for Random MAX-2-XORSAT. LATIN Int. Conf., 2010, pages 320-331.

3.4 Exposés et articles invités

N. Creignou a été invitée à donner deux conférences en relation avec ses travaux sur les instances aléatoires : Workshop on tractability, Microsoft Research, Cambridge (U.K.), 6-8 juillet 2010 ; Guangzhou Symposium on Satisfiability in Logic-Based Modeling, Zhuhai, Chine, 25-29 Septembre 2010.

C. Carlet a été invité à écrire un article sur le profil de non-linéarité de la fonction courbe $f(x, y) = x/y$, pour un numéro spécial du journal Foundations of Computer Science.