Vote par sondage uniforme incorruptible

Nicolas Blanchard

Vote par sondage uniforme incorruptible

Nicolas K. Blanchard[1]

[1]*ENS Paris, IRIF,* `http://www.liafa.univ-paris-diderot.fr/~nkblanchard`

Introduit en 2012 par David Chaum, le vote par sondage uniforme (*random-sample voting*) est un protocole de vote basé sur un choix d'une sous-population représentative, permettant de limiter les coûts tout en ayant de nombreux avantages, principalement lorsqu'il est couplé à d'autres techniques comme ThreeBallot. Nous analysons un problème de corruptibilité potentielle où les votants peuvent vendre leur vote au plus offrant et proposons une variation du protocole remédiant à ce problème, ainsi que plusieurs extensions possibles selon les propriétés désirées.

Keywords: Random sample voting, Decision theory, Online voting, Probabilistic protocol

1 Introduction

Dès les toutes premières démocraties il y a 2500 ans, le hasard était utilisé pour garantir une égalité entre citoyens et limiter la corruption lors du choix des gouvernants (qui étaient choisis au hasard parmi l'assemblée). Ce système peut cependant sélectionner des personnes incompétentes, et il fut donc presque abandonné dans les démocraties modernes, subsistant principalement dans les jurys criminels. L'idée d'un hasard représentatif a toutefois de nombreux avantages, démontrés depuis quelques années par les équipes de James Fishkin et David Chaum et le RSVP [†]. Le vote par sondage uniforme – proposé par ce dernier [Cha14, CP] – est un protocole transparent qui sélectionne une fraction de la population de manière publiquement vérifiable et ne compte que leurs votes pour l'élection finale. Les bénéfices sont nombreux, tant économiques que politiques. Le protocole initialement proposé a toutefois un problème : il est relativement facile pour un électeur de vendre son vote au plus offrant. Nous proposons et analysons ici plusieurs variantes du protocole résolvant ce problème. Couplés à des outils cryptographiques et à Threeballot [*] [RS07], ces protocoles pourraient améliorer la sécurité et l'efficacité des systèmes de vote en ligne qui commencent à être utilisés, notamment en Estonie [SFD+14], mais qui sont pour le moment assez vulnérables.

2.1 Protocoles de vote

Le protocole initialement proposé par le RSVP est remarquablement simple, présentons le d'abord pour une élection simple :

1. $n = 10000$ personnes sont choisies uniformément au hasard, de manière vérifiable dans la population.
2. Elles reçoivent un bulletin de vote avec un identifiant unique, et l'envoient au comité électoral.
3. Les bulletins sont analysés, et seuls ceux ayant les bons identifiants sont décomptés.

†. Random Sample Voting Project (rsvoting.org) organisation dont l'auteur fait désormais partie.

∗. Dans ce système, chacun des n votants envoie $k+1$ bulletins de vote, un pour chacun des k candidats ainsi qu'un dernier pour celui qu'il choisit. Cela rajoute n votes à chaque candidat ce qui ne change pas le résultat. Chaque bulletin a de plus un identifiant unique, et le votant conserve un seul des $k+1$ identifiants au hasard. On publie alors tous les bulletins et procède au décompte. Si un bulletin est erroné, le votant concerné a une probabilité $1/(k+1)$ de se rendre compte de l'erreur. Modifier plus que quelques votes est donc systématiquement visible. De plus, personne ne peut savoir pour qui a voté chaque autre votant car chacun a voté pour tous les candidats et n'a conservé qu'un seul identifiant, garantissant ainsi l'anonymat.

Ici on part du principe que chaque personne a un identifiant unique privé, et que le choix des votants se fait à partir d'un hasard publiquement accessible (tirage de numéros à la télévision ou phénomène boursier difficile à contrôler et ayant une probabilité quasi uniforme de choisir chaque participant). Ainsi chacun peut vérifier qu'il avait bien une probabilité correcte d'être choisi.

Ce système a de nombreux avantages (détaillés plus bas), mais a une faille : il est possible de vendre son vote, pour un montant élevé vu la rareté des votes, et si les politiciens ne peuvent pas savoir à qui s'adresser, il suffit qu'un des votants prenne l'initiative pour avoir un système corruptible. Dans l'article original, les auteurs proposaient d'envoyer un certain nombre de faux bulletins, pour diluer cette valeur, mais cela pose problème. Si les votants ne peuvent pas faire la différence le public perd confiance envers le système. S'ils peuvent faire la différence, cela rend le système plus complexe et réduit encore la compréhension du public. Nous proposons donc le protocole suivant, visant à empêcher cette possibilité de corruption :

1. Un groupe de n individus est choisi au hasard uniformément parmi l'ensemble de la population. Chacun de ces individus reçoit un courrier contenant un identifiant unique et un mot de passe.
2. Les n individus choisis ont tous accès à un système d'identification grâce auquel ils peuvent enregistrer leur vote (pendant une période limitée), avec une date limite 1 à 2 mois après la réception du courrier.
3. Chaque citoyen a aussi accès à une copie du courrier envoyé et à un système permettant de générer des faux identifiants. Ces identifiants sont indistinguables des identifiants réels mais les votes correspondants ne sont pas comptabilisés dans le total final.
4. Le vote en ligne utilise ThreeBallot et publie tous les bulletins ainsi que le décompte après le vote.
5. Chaque votant peut alors vérifier que son vote a bien été comptabilisé dans le total.

Si le principe est identique, il y a quelques différences notables. Tout d'abord on repose dès le départ sur un système (Identifiant, Mot de passe), créé de manière cryptographique pour être sûr que générer des paires valides – ou trouver le Mot de passe pour un identifiant – est difficile d'un point de vue algorithmique. Ensuite l'utilisation de ThreeBallot [RS07] permet de rajouter une étape de vérification supplémentaire garantissant que même dans le cas de modification des données les utilisateurs peuvent vérifier l'intégrité du système. Ce protocole a encore plusieurs défauts, mais on peut déjà observer certaines propriétés désirables :

2.2 Avantages

1. (Représentativité) Tout d'abord, un élément crucial est que, en calculant de manière appropriée le nombre de votants nécessaires, l'échantillon est représentatif. Une représentativité parfaite est naturellement impossible, mais si le décompte est fait de manière correcte – ce qui est automatisé et vérifiable – on peut rapidement arriver à un résultat suffisament précis. En effet, avoir 5000 votants suffit pour garantir que l'erreur est inférieure à 2.5% dans 99.9% des votes ; et avec 150000 votants on passe à une marge d'erreur inférieure à 0.5%. Ce n'est pas idéal, mais les données sur les élections américaines [GBG12] – les données françaises n'étant pas disponibles – suggèrent que la marge d'erreur dans un décompte varie entre 0.5% et 2%.
2. (Vérifiabilité) L'utilisation du système ThreeBallot et de numéros tirés au hasard publiquement permet de déceler les manipulations à chaque étape, et tout citoyen peut le vérifier. De plus, la coercicion devient presque impossible, étant donné qu'il est impossible de savoir qui a voté, et pour qui.
3. (Motivation) Vu que chaque votant a une réelle importance (la dilution de responsabilité n'ayant plus lieu, et chaque vote ayant un réel impact), le taux d'absentéisme parmi ceux devant voter devrait être bas, et le temps passé à comparer les campagnes devrait augmenter. Cela a une deuxième conséquence importante : les votants passant plus de temps à étudier les différentes possibilités, l'impact de la publicité sera beaucoup plus faible, permettant de diminuer l'impact de celle-ci – et donc des besoins importants de financement pour les campagnes.

4. (Incorruptibilité) Vu que n'importe quel individu peut créer un faux bulletin de vote, il suffit qu'une personne sur mille essaie de vendre un faux bulletin si un politicien est corruptible pour que le marché soit inondé de vendeurs et que ça ne vaille pas le coup d'essayer d'acheter des voix.
5. (Coût) Ce protocole étant informatisé, il coûterait moins que le système actuel.

2.3 Critiques potentielles et solutions

Malgré les avantages précédents, plusieurs problèmes subsistent dans ce système, que l'on peut toutefois mitiger.

1. (Vulnérabilité informatique) Tout système reposant sur un vote électronique a inhéremment des failles potentielles [SFD+14]. Le fait que le choix des votants se fasse avec du hasard public et que les votes soient aussi décomptés publiquement et visibles par tous via le système ThreeBallot limite grandement l'impact potentiel du piratage.
2. (Impopularité) Ce protocole utilisant du hasard, il paraît contre-intuitif et injuste. Une grande partie de la population sera probablement contre par défaut, ce qui serait un obstacle à son implémentation *.
3. (Possibilité d'erreur) Comme ce protocole est probabiliste par nature, il a une probabilité intrinsèque d'erreur. On peut cependant contrer cela avec un système adaptatif : on choisit une probabilité d'erreur p suffisament faible (par exemple $p = 0.001$) et on fait le sondage avec $n = 5000$ votants. Si on a une marge entre les candidats telle qu'il y ait une probabilité supérieure à p d'avoir une erreur dans le gagnant, on recommence avec $n = 30000$ puis $n = 150000$, permettant de départager même les élections très équilibrées †.
4. (Manipulation physique) Envoyer les identifiants par courrier pourrait être traçable et dangereux, mais peut être limité par l'envoi de tels courriers avec un courrier déjà distribué à tous (comme les impôts jusqu'à récemment). Une solution idéale serait d'utiliser un système d'identité virtuelle sécurisée comme celle utilisée aujourd'hui en Estonie [SFD+14].
5. (Preuve d'identité) Vu qu'il n'y a ni isoloir ni vérification d'une pièce d'identité, un autre individu pourrait utiliser l'identifiant, par exemple un membre de sa famille. L'utilisation d'une pièce d'identité virtuelle limiterait aussi l'ampleur de ce problème.
6. (Perturbation) Si chacun peut générer des faux identifiants on pourrait imaginer un individu imprimant en grand nombre des enveloppes et les envoyant à des citoyens aléatoires pour perturber l'élection et diminuer la confiance en le système. Pour empêcher cela il est possible d'utiliser une méthode cryptographique, mais il y a une solution plus simple : mettre 100€ dans l'enveloppe, ce qui rendrait une perturbation très coûteuse.
7. (Manque de vérifieurs) Pour le choix des votants et la vérification du vote une association de citoyens pourrait faire l'affaire, mais l'impression de faux bulletins par des individus pour noyer le marché est potentiellement instable du point de vue théorie des jeux. Une manière de compenser cela serait d'envoyer tout de même une fraction de faux bulletins en envoyant un deuxième courrier aux mêmes personnes les informant que leur bulletin est faux, et d'envoyer de l'argent comme précédemment pour compenser l'incompréhension et la perte de confiance envers le système.

Ce système a encore des vulnérabilités mais a déjà un intérêt plus que théorique et les premiers tests pratiques en Oregon ont eu du succès [Zha]. Naturellement, une généralisation et application à un pays entier n'est pas dans l'horizon immédiat, mais en ayant une optique de long-terme, on peut étudier une évolution de ce protocole comme possibilité pour un futur système démocratique équitable et efficace.

∗. La vérifiabilité par tout le monde du système a été testé par le RSVP en Oregon, et semble compenser l'à-priori négatif.

†. Le record pour une présidentielle en France était de 1,62% d'écart en 1974.

3. Démo-techno-cratie

Si l'idée de technocratie – le gouvernement par les experts – paraît tentante, elle tend en pratique à donner le pouvoir à une classe endogame ayant intérêt et les moyens de garder le reste de la population dans l'ignorance. Comme on ne peut faire voter uniquement les experts, l'autre idée est d'expertiser les votants. Le système présenté auparavant peut en effet être très facilement étendu pour former une démocratie directe où la population est représentée à chaque choix, sans pourtant que chaque citoyen passe un temps considérable à étudier chaque proposition. Plusieurs extensions permettent en effet de réaliser cela :

1. Tout d'abord, le fait d'avoir une élection avec un décompte public en ligne fait que l'élection peut s'étendre dans le temps, et que les votants peuvent être prévenus plusieurs mois à l'avance.
2. La présence d'un débat public et accessible à tous, à travers l'utilisation d'un système inspiré par exemple de https://www.republique-numerique.fr permettrait de garantir un accès de tous à l'information et à l'avis populaire sur les différents points de loi.
3. Une rémunération financière garantie, par l'envoi de 100€ avec les identifiants qui pourrait compenser l'investissement temporel et augmenter la motivation [*]. Même à raison d'un vote par jour, et d'une fraction de votes répétés en cas de quasi-égalité, on arrive à un peu moins de 300M€ par an avec 100€ par enveloppe (nos élections actuelles coutant 350M€ par an).

Ce système possède lui aussi des défauts, principalement par le fait que le pouvoir se trouverait principalement dans les mains de ceux qui formulent les lois sur lesquelles les citoyens votent (ce qui peut être mitigé par le fait d'avoir un forum public et des législateurs élus eux aussi). Malgré les défauts cette classe de protocole est riche et peut encore être raffinée, et ses effets politiques méritent d'être étudiés en détails (les premières implémentations pratiques sont très récentes – testées à CRYPTO 2015 et RWC 2016 – et aucune étude politique ou sociologique n'a pu avoir lieu pour le moment). À ce jour ces protocoles semblent déjà être comparables voire meilleurs que la plupart des systèmes actuels, et l'ajout de couches cryptographiques pourrait réduire à néant plusieurs objections restantes.

Références

[Cha14] David Chaum. Random sample elections, June 19 2014. US Patent App. 14/237,991.

[CP] David Chaum and Random Sample Voting Project. Random-sample voting, white paper. Updated monthly at `http://rsvoting.org/whitepaper/white_paper.pdf`.

[GBG12] Stephen N. Goggin, Michael D. Byrne, and Juan E. Gilbert. Post-election auditing : Effects of procedure and ballot type on manual counting accuracy, efficiency, and auditor satisfaction and confidence. *Election Law Journal : Rules, Politics, and Policy*, 11(1) :36–51, 2012.

[Kam12] Emir Kamenica. Behavioral economics and psychology of incentives. *Annual Review of Economics*, 4 :427–452, 2012.

[RS07] Ronald L. Rivest and Warren D. Smith. Three voting protocols : Threeballot, vav, and twin. In *Proceedings of the USENIX Workshop on Accurate Electronic Voting Technology*, EVT'07, pages 16–16, Berkeley, CA, USA, 2007. USENIX Association.

[SFD+14] Drew Springall, Travis Finkenauer, Zakir Durumeric, Jason Kitcat, Harri Hursti, Margaret MacAlpine, and J. Alex Halderman. Security analysis of the estonian internet voting system. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, pages 703–715, New York, NY, USA, 2014. ACM.

[Zha] Bingsheng Zhang. Random sample voting implementation technical report.

∗. Les effets sur la motivation sont à étudier plus précisement, voir [Kam12].