

Rumeurs, populations et communautés : équivalence uniquement sociologique ? Protocole de population versus protocoles épidémiques

Marin Bertier, Yann Busnel, Anne-Marie Kermarrec

► To cite this version:

Marin Bertier, Yann Busnel, Anne-Marie Kermarrec. Rumeurs, populations et communautés : équivalence uniquement sociologique ? Protocole de population versus protocoles épidémiques. Maria Gradinariu Potop-Butucaru and Hervé Rivano. 12èmes Rencontres Francophones sur les Aspects Algorithmiques de Télécommunications (AlgoTel), 2010, Belle Dune, France. 2010. <inria-00475746>

HAL Id: inria-00475746

<https://hal.inria.fr/inria-00475746>

Submitted on 22 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Rumeurs, populations et communautés : équivalence uniquement sociologique ?

Protocole de population versus protocoles épidémiques

Marin Bertier¹, Yann Busnel² et Anne-Marie Kermarrec³

¹ IRISA / INSA Rennes, Campus universitaire de Beaulieu, F-35042 Rennes Cedex, France

² LINA / Université de Nantes, 2 rue de la Houssinière, BP 92208, F-44322 Nantes Cedex 03, France

³ INRIA Rennes – Bretagne Atlantique, Campus universitaire de Beaulieu, F-35042 Rennes Cedex, France

Les protocoles épidémiques sont simples, robustes, passent à l'échelle et ont déjà été appliqués à de nombreux systèmes distribués (principalement filaires). Cependant, la majorité des validations de ce domaine ont été conduites de manière empirique et il persiste un manque significatif de caractérisation théorique des possibilités de ces protocoles épidémiques. D'un autre côté, les protocoles de population modélisant les réseaux mobiles bénéficient d'un cadre théorique solide mais de très peu d'évaluations empiriques. Dans cet article, nous établissons une corrélation entre les protocoles épidémiques et de population. Pour cela, nous proposons une classification des protocoles épidémiques, dépendants du service d'échantillonnage sous-jacent. Nous montrons dans un premier temps que la classe des protocoles épidémiques fondés sur des échantillons arbitraires est équivalente aux protocoles de population. Puis, nous montrons que les protocoles épidémiques reposant sur un service d'échantillonnage plus puissant (par l'identification explicite des nœuds) sont équivalents aux protocoles de communauté, une extension récente des protocoles de population. Il est ainsi possible de tirer profit de ces équivalences : (i) par la définition d'un cadre théorique pour les systèmes distribués où un comportement global émerge d'interactions locales, à la fois dans un contexte filaire et sans-fil et (ii) par le transfert des connaissances empiriques de ces derniers systèmes dans les réseaux mobiles modélisés par les protocoles de population. En sus de l'application aux résultats existants, ces équivalences ouvrent des opportunités de transfert de compétences d'un domaine à l'autre, à la fois théorique et pratique.

Keywords: Protocole de population; Protocole épidémique; Equivalence de modèles; Classification

1 Introduction et contexte

Les applications et résultats de recherche concernant les réseaux dynamiques et/ou mobiles sont légion depuis ces dernières années. Cependant, ceux-ci se rapportent le plus souvent à des aspects pratiques spécifiques des systèmes mobiles, ne permettant de valider le plus souvent qu'empiriquement les solutions proposées.

Protocoles de population et de communauté Les protocoles de population (PP), initialement introduits dans [AAD⁺06], ont été conçus originellement pour représenter les réseaux de capteurs composés d'agents mobiles à très faibles ressources, n'ayant aucun contrôle sur leur propre mouvement. Ainsi, ceux-ci modélisent les interactions dans un ensemble donné d'agents. Ce modèle fournit une base théorique commune pour les systèmes répartis, sur lesquels un comportement global émerge d'un ensemble d'interactions locales entre les nœuds [AR07]. Un tel modèle peut être utilisé dans de nombreux contextes tels que les réseaux mobiles ad-hoc (MANET) ou les réseaux de capteurs sans fil.

L'abstraction des PP consiste en un ensemble fini d'états, un ensemble fini d'entrées et de sorties ainsi qu'une fonction de transition. L'ensemble des interactions possibles est représenté sous la forme d'un graphe d'interactions. Si deux agents mobiles se trouvent être suffisamment proches l'un de l'autre durant un laps de temps suffisamment long, ils interagissent en échangeant leurs informations locales. Par exemple, en considérant de petites entités embarquées sur des animaux, une interaction a lieu chaque fois qu'un animal équipé se trouve à proximité du rayon de transception d'un autre.

La puissance des PP repose sur la simplicité du modèle. Les agents sont considérés comme anonymes et aucune hypothèse spécifique n'est faite sur le synchronisme, l'ordre des interactions, l'infrastructure émergente ou même la taille du système. La suite d'interactions est uniquement guidée par un *ordonnanceur* sur lequel une unique hypothèse est imposée : celle d'être *équitable* (i.e. de s'assurer que toute évolution possible du système ne pourra être évitée indéfiniment).

En marge, de nombreuses extensions ont été proposées parmi lesquelles les *protocoles de communauté* (PC) [GR07]. L'anonymat des agents dans les PP contraste fortement avec la grande majorité des modèles de systèmes répartis, au sein desquels la possession d'un identifiant unique par processus est souvent une propriété cruciale des algorithmes. Ainsi, récemment, Guerraoui et Ruppert ont proposé ce modèle étendu qui, tout en conservant l'essence des PP par la nature minimaliste des agents, permet à ces derniers de se voir attribuer des identifiants uniques issus d'un ensemble déterminé. Néanmoins, afin d'éviter une extension du modèle dans laquelle les espaces mémoires dédiés au stockage d'identifiants soient utilisés pour conserver une masse d'information arbitrairement grande, l'utilisation des identifiants est limité à l'*identification*. Ceci permet de généraliser les PP, tout en gardant l'esprit minimaliste de ceux-ci, en interdisant aux agents d'utiliser l'espace réservé au stockage des identifiants pour accroître leur puissance de calcul.

Protocoles épidémiques Par ailleurs, en analogie avec le cheminement d'une rumeur ou la propagation d'un virus, les *protocoles épidémiques* sont conçus autour d'un échange périodique d'informations entre les participants d'un système réparti, lequel peut tout aussi bien être filaire ou non. Dans le contexte des systèmes filaires, ces protocoles fournissent une trame fiable et robuste, même à grande échelle, pouvant être utilisée dans une large gamme d'application [KvS07], telles que la cohérence de bases de données, la diffusion d'informations, la construction de réseaux logiques, etc. La robustesse des protocoles épidémiques reposent sur leur caractère aléatoire : dans cette classe de protocoles, chaque nœud du système échange périodiquement des informations avec un autre nœud choisi au hasard parmi un sous-ensemble connu du système. La taille de cet ensemble est classiquement d'un ordre de grandeur bien plus faible que celui de la taille du système. Le choix de ce sous-ensemble, la plupart du temps dénoté *vue de voisinage*, est cruciale pour l'efficacité de la dissémination épidémique [JVG⁺07].

A notre connaissance, aucun modèle formel de protocoles épidémiques n'a encore été proposé. La majorité des résultats de cette classe de protocoles est observée empiriquement, bien que quelques études théoriques aient déjà permis d'extraire certaines propriétés. De plus, aucune classification des différents protocoles existants n'a encore été proposée à ce jour. Proposer une telle classification des protocoles épidémiques est un des objectifs de cet article, inspirée des classes issues des modèles de PP / PC et fondée sur la puissance de l'échantillonnage des nœuds sous-jacent (ou *PS* pour *Peer Sampling*).

2 Présentation des contributions

Les protocoles de population et les protocoles épidémiques possèdent de nombreuses caractéristiques communes. Premièrement, les deux reposent sur une séquence d'interactions menée par un *ordonnanceur*. Dans le contexte des PP, ce dernier spécifie les interactions des agents de manière équitable dans un environnement mobile. Au sein des protocoles épidémiques, l'ordonnanceur fournit un service d'échantillonnage des nœuds lequel approvisionne les nœuds participants aux échanges. De plus, l'objectif de chacun de ces protocoles est de faire émerger un comportement global à partir d'un ensemble d'interactions locales, de manière totalement décentralisée.

Bien qu'ils aient été étudiés par deux communautés différentes, d'autres ressemblances significatives existent entre ces deux modèles. Chacune de ces classes de protocoles repose sur les propriétés suivantes :

- Un modèle complètement décentralisé ;
- Un ensemble d'agents possédant une capacité de stockage finie et interagissant deux-à-deux. Les agents sont mobiles à communication sans fil dans les PP. Dans les protocoles épidémiques, ils sont statiques mais communiquent via un réseau dynamique sur une infrastructure fixe ;
- Une séquence d'interactions dirigée par un ordonnanceur équitable modélisant la mobilité des agents dans les PP, ou par un service d'échantillonnage de nœuds (PS) dans les protocoles épidémiques ;
- Une fonction spécifiant le procédé suivant lequel les données sont traitées au cours d'une interaction : ceci correspond à la fonction de transition des PP et à la fonction de mise à jour des protocoles

épidémiques ;

- Un échange d'états durant l'interaction, lesquels correspondent à une valeur dans l'espace des états pour les PP et au retour de la fonction de sélection d'information dans les protocoles épidémiques.

Dans cet article, nous résumons certains résultats démontrés dans [BBK09], correspondant notamment aux contributions suivantes :

- Nous établissons une corrélation entre les PP en général et les protocoles épidémiques, par la définition d'une *classification* des protocoles épidémiques. Nous identifions ainsi quatre classes de protocoles épidémiques (cf. partie gauche de la figure 1) ;
- Nous montrons que les protocoles épidémiques asynchrones sur des nœuds anonymes sont *équivalents* au modèle initial des PP ;
- Nous montrons que les protocoles épidémiques à nœuds avec identifiants sont *équivalents* aux PC ;

Grâce à ces équivalences, nous pouvons tirer profit du cadre théorique des PP pour comprendre et analyser la puissance et les limitations des protocoles épidémiques. De même, nous pouvons exploiter les résultats obtenus dans le domaine des protocoles épidémiques pour tirer des conclusions quant à l'utilisation pratique des PP. Ceci permet de fournir des considérations à la fois théoriques et pratiques pour ces systèmes large-échelle : ce parallèle entre PP et épidémique peut ainsi être exploité pour des résultats existants et à venir.

En raison de la contrainte d'espace, nous ne pouvons développer plus en détails chacun de ces modèles. Pour plus d'information sur la définition et la puissance de ceux-ci, le lecteur pourra se référer à [BBK09]. Nous présentons ci-après les critères de classification des protocoles épidémiques, résumés en figure 1 puis leurs relations d'équivalences avec les modèles de population et de communauté.

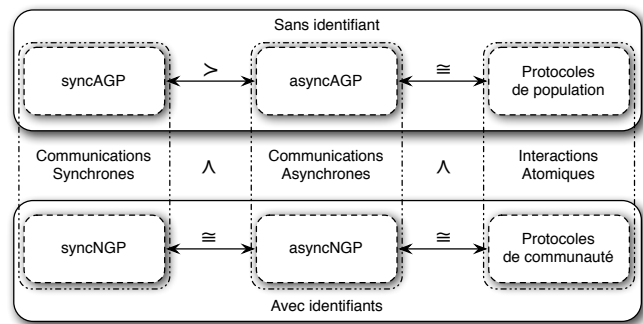


FIG. 1: Relations entre protocoles épidémiques et de population

3 Classification des protocoles épidémiques

Un parallèle peut être établi entre (i) les différences entre PP et PC avec (ii) la nécessité d'identification des nœuds dans les protocoles épidémiques.

En effet, les protocoles épidémiques diffèrent selon les besoins d'anonymat, afférent au service d'échantillonnage sous-jacent. Ce dernier peut être vu comme une "boîte noire" fournissant un nœud issu d'un échantillon quelconque du réseau au protocole épidémique. Deux classes de ces protocoles sont ainsi définies selon l'anonymat ou non des nœuds fournis par le service d'échantillonnage.

AGP : Anonymous Gossip protocols Classe des protocoles épidémiques ne nécessitant pas la connaissance de l'identifiant d'aucun nœud pour s'exécuter. C'est typiquement le cas des protocoles de calculs distribués tels que l'estimation de la taille de système ou les calculs de moyenne, ainsi que les protocoles de diffusion à l'aveugle.

NGP : Non-anonymous Gossip Protocols Classe de protocoles épidémiques dont la connaissance de l'identifiant du nœud distant, partenaire de l'échange périodique, est nécessaire. Typiquement, les protocoles de construction de réseau logique, de classement de nœuds, *etc.* appartiennent à cette classe.

Dès à présent, nous pouvons déduire de manière triviale que $AGP \prec NGP$.

Entre synchronisme et asynchronisme

Afin d'établir une équivalence avec les PP, il est nécessaire de prendre en compte les caractéristiques des canaux de communications entre les nœuds, afin de raffiner la classification des protocoles épidémiques.

Dans le modèle asynchrone, le délai de transmission des messages, décalage d'horloge, temps d'exécution sont bornés et les bornes sont connues. À l'inverse, le modèle asynchrone n'implique aucune borne. Dans

ce dernier modèle, il n'est donc pas possible de tirer avantage de la périodicité des protocoles épidémiques.

Classification finale

La prise en compte du synchronisme nous permet d'affiner le grain de la classification sus-présentée :

- syncAGP** Synchronous Communication and Anonymous Nodes ;
- asyncAGP** Asynchronous Communication and Anonymous Nodes ;
- syncNGP** Synchronous Communication and Non-anonymous Nodes ;
- asyncNGP** Asynchronous Communication and Non-anonymous Nodes.

4 Comblant le fossé entre les deux domaines

Le synchronisme permettant de tirer profit de la périodicité des protocoles épidémiques, la puissance de calcul est donc augmentée. Il est donc trivial que $\text{asyncAGP} \prec \text{syncAGP}$ et $\text{asyncNGP} \prec \text{syncNGP}$. De surcroît, nous prouvons qu'il existe une équivalence entre les protocoles épidémiques asynchrone et les protocoles de population/communauté, soit $\text{asyncAGP} \cong \text{PP}$ et $\text{asyncNGP} \cong \text{PC}$. La preuve de ces équivalences repose sur le fait qu'il est possible de simuler tout protocole d'une de ces classes par un protocole de l'autre classe et *vice-versa*. Les détails de ces preuves sont disponible dans [BBK09, Th. 1 & 2].

Enfin, la preuve du Théorème 2 de [BBK09] montre qu'il est possible de simuler la périodicité en milieu asynchrone, via l'utilisation de barrières de synchronisation. Ce dernier résultat permet donc de conclure que $\text{syncNGP} \prec \text{asyncNGP}$ et par conséquent que $\text{syncNGP} \cong \text{asyncNGP}$.

L'ensemble de ces relations, ouvrant des perspectives attractives, sont résumées en figure 1.

5 Un impact potentiel pour les deux domaines

Les équivalences en amont sont de la plus grande importance dans le domaine des protocoles épidémiques. Elles permettent notamment de définir clairement ce qui peut être calculable par un protocole épidémique anonyme, *i.e.* toute fonction issue de l'arithmétique de Presburger peut ainsi être calculable par un membre de la classe AGP. Nous pouvons également conclure qu'aucune autre fonction ne peut être calculable par un protocole de asyncAGP [AAD⁺06, AAE06]. Par ailleurs, les résultats empiriques sur la vitesse de convergence et la mise en pratique des protocoles épidémiques peuvent être immédiatement appliqués afin d'évaluer l'efficacité des protocoles de population et de communauté.

De même, les protocoles épidémiques reposant sur un ensemble de nœuds identifiables sont équivalents aux PC. Ceci permet de conclure que l'ensemble des fonctions calculables par les protocoles de la classe NGP correspondent à l'ensemble des fonctions symétriques de $NSPACE(n \log n)$, et peuvent également mettre en œuvre des algorithmes tolérants à des défaillances bénignes ou non. L'importance de ces résultats est de pouvoir tirer profit des résultats existants, et à venir, pour chacun des deux domaines.

Références

- [AAD⁺06] D. Angluin, J. Aspnes, Z. Diamadi, M.J. Fischer, and R. Peralta. Computation in networks of passively mobile finite-state sensors. *Distributed Computing (Special Issue : PODC'04)*, 18(4) : 235–253, March 2006.
- [AAE06] D. Angluin, J. Aspnes, and D. Eisenstat. Stably computable predicates are semilinear. In *25th annual ACM Symposium on Principles of Distributed Computing (PODC '06)*, pages 292–299, August 2006.
- [AR07] J. Aspnes and E. Ruppert. An introduction to population protocols. *Bulletin of the European Association for Theoretical Computer Science, Distributed Computing Column*, 93 : 98–117, October 2007.
- [BBK09] M. Bertier, Y. Busnel, and A.-M. Kermarrec. On Gossip and Population. In *the 16th International Colloquium on Structural Information and Communication Complexity (SIROCCO '09)*, Piran, Slovenia, May 2009.
- [GR07] R. Guerraoui and E. Ruppert. Even small birds are unique : Population protocols with identifiers. Technical Report Technical Report CSE-2007-04, Dept of Computer Science and Engineering, York University, Sep, 2007.
- [JVG⁺07] M. Jelasity, S. Voulgaris, R. Guerraoui, A.-M. Kermarrec, and M. van Steen. Gossip-based peer sampling. *ACM Transactions on Computer Systems*, 25(3) : 8, 2007.
- [KvS07] A.-M. Kermarrec and M. van Steen, ed. *ACM Operating Systems Review on Gossip Potocols*, volume 41(5), October 2007.