Centralisé, décentralisé, pair à pair, quels mots pour l'architecture des systèmes répartis ?

Stéphane Bortzmeyer
AFNIC
1, rue Stephenson
78 180 Montigny-le-Bretonneux

Résumé

Dans le milieu des acteurs de l'Internet, il y a depuis des années des débats autour de l'architecture des systèmes répartis. Par exemple, on va critiquer le fait que Facebook soit « un système centralisé » et on va chercher à faire un équivalent qui n'ait pas cette propriété. Ou bien on va se féliciter que BitTorrent soit « pair à pair ». Les révélations de Snowden en 2013 ont évidemment intensifié ces débats, les fameux GAFA [1] étant, non seulement centralisés, mais également fournisseurs du système d'espionnage PRISM [2].

Le problème de ces débats est que les mots sont souvent utilisés de manière très laxiste, voire à contre-sens. On entend ainsi des gens critiquer le fait que le Domain Name System (DNS) soit centralisé (ce qui est faux) tout en se félicitant d'utiliser Internet Relay Chat (IRC) (qui est, lui, réellement centralisé). Et que dire des dépendances entre systèmes ? BitTorrent est-il vraiment pair à pair, dans l'utilisation qu'en font les gens (avec un moteur de recherche comme ThePirateBay) ?

On va essayer de définir différemment les différentes classes de systèmes répartis, selon le rôle que peut jouer une organisation particulière. Dans la classe 1, une entité peut décider de tout, dans la classe 2, une délégation permet à d'autres entités d'agir de manière autonome, mais avec toujours le risque d'être « déconnectée », dans la classe 3 on dépend d'un accord des pairs pour pouvoir participer, dans la classe 4, chacun est « réellement » autonome.

Mots clefs

P2P, pair à pair, centralisation, décentralisation, architecture, terminologie

1 Introduction

Il y a depuis de nombreuses années des débats autour de l'architecture des systèmes répartis. En général, l'idée dominante chez les « numéristes [3] » est qu'un système « décentralisé » est bon et qu'un « centralisé » est mauvais. En effet, dans un système centralisé, on est à la merci d'une entité unique, individu, entreprise ou gouvernement, qui peut décider du jour au lendemain de vous couper l'accès. Par exemple, on va critiquer le fait que Facebook soit « un système centralisé ». Cette centralisation permet à Facebook de supprimer le contenu qui le dérange (une décapitation, c'est OK, un sein nu, c'est intolérable). De même, le système

1. *Google, Apple, Facebook et Amazon*
2. https://fr.wikipedia.org/wiki/PRISM_(programme_de_surveillance)
3. Merci à B. Cazeneuve pour cet excellent terme, qu'il voulait insultant.

de partage de fichiers Napster a pu être stoppé efficacement par l'industrie du divertissement car son index était centralisé.

Les révélations d'Edward Snowden en 2013 ont évidemment intensifié ces débats, les fameux GAFA, les gros acteurs du Web, les seuls connus pour pas mal de dirigeants politiques, étant, non seulement centralisés, mais également fournisseurs de la National Security Agency (NSA) en données.

Vu cette idée dominante, beaucoup de gens vont chercher à faire un équivalent décentralisé des systèmes centralisés (C'était le projet explicite de Diaspora [4], porté à l'origine par des gens qui n'avaient jamais étudié les systèmes décentralisés existants et qui prétendaient être les premiers sur ce terrain.)

Le problème est que les termes utilisés, « centralisé », « acentré », « décentralisé », « pair à pair », ne sont jamais définis précisément et, quand ils le sont, chacun a sa propre définition. Ils sont donc plus souvent utilisés comme anathèmes dans le débat (« le DNS, c'est mal, car c'est centralisé ») que comme vraie terminologie technique.

Un exemple de cette confusion m'avait été fourni par un fana d'IRC qui prétendait qu'IRC était décentralisé car un serveur IRC est mis en œuvre sur plusieurs machines physiques différentes. La plupart des techniciens rejetteraient cette interprétation (en effet, à ce compte, Gmail est lui aussi décentralisé !) mais elle a été utilisée par des organismes importants.

Autre exemple de confusion, lorsque les gens n'analysent qu'une partie d'un système. Ainsi, on entend souvent dire qu'en échangeant des contenus culturels via BitTorrent, on est en « pair à pair ». Mais cela ne reflète pas la réalité de l'expérience de la plupart des gens, qui doivent d'abord récupérer un *magnet* ou autre indicateur via un moteur de recherche très centralisé, comme ThePirateBay ou T411 [5].

Plutôt que d'essayer de redresser le vocabulaire en donnant « la bonne » définition de termes comme centralisé ou décentralisé, avant de me battre pour faire reconnaitre cette bonne définition, je vais plutôt présenter ici une terminologie nouvelle, qui a l'avantage de ne pas reposer sur les idées existantes. Je vois quatre grandes catégories ou *classes*. On va les décrire en fonction de *qui* peut modifier (sans autorisation) l'*état* d'un système.

2 Classe 1

La classe 1 regroupe les systèmes où toute modification de l'état nécessite l'accord d'une entité donnée. Par exemple, Gmail est un système de classe 1, tout ce qui s'y passe nécessite l'accord de Google. À noter que cette définition ne se soucie pas de savoir si le système fonctionne sur une seule machine ou sur plusieurs [6], un point qui est vu comme juste un détail de mise en œuvre [7].

Quand je parle d'accord, cela ne signifie pas forcément un accord explicite et signé d'un humain. Cet accord peut être plus ou moins implicite, par exemple via le code [8] Ainsi, le logiciel de Facebook décide de ce que l'utilisateur va voir sur son « mur » [9]. Ses critères sont opaques et toujours changeants mais le point important est que l'utilisateur (ou bien la marque qui voudrait que ses messages sans intérêt soient toujours affichés…) n'a guère de marge de manœuvre : Facebook décide.

Cette entité, cette organisation qui dirige le système de classe 1 peut donc développer une politique (interdisant par exemple certaines activités) et la faire respecter. Et si cette organisation a une défaillance, le système cesse de fonctionner. Les GAFA sont tous des systèmes de classe 1 [10]. Par exemple, pour Gmail,

4. https://fr.wikipedia.org/wiki/Diaspora*
5. Il y a des moteurs de recherche non centralisés comme Tribler http://www.tribler.org/ mais très peu répandus.
6. Avec la virtualisation, ce point est de toute façon de plus en plus dur à déterminer. Une machine physique dont tous les disques sont externes et accessibles en iSCSI est-elle une machine unique ou non ?
7. Même si ce « détail » va sérieusement compliquer la vie du programmeur !
8. « *The code is the law* », selon la fameuse phrase de Lawrence Lessig [1].
9. http://pro.clubic.com/blog-forum-reseaux-sociaux/facebook/actualite-764198-facebook-algorithme.html
10. On n'a pas cité Twitter, mais il a les mêmes caractéristiques, d'où le développement de son concurrent Twister [2].

le choix de l'identificateur, la détermination de qui aura tel compte, le choix de la politique anti-spam sont tous contrôlés par Google et Google seul.

Dans un genre plus historique, l'ancien `HOSTS.TXT` d'Internic était également un système de classe 1. Longtemps avant le DNS, toutes les machines de l'Internet étaient enregistrées dans une base de données centrale, maintenue par Elizabeth Feinler [3]. Tout changement d'adresse d'une machine quelconque de l'Internet nécessitait donc d'envoyer un message à Stanford et d'attendre que l'employée de l'Internic ait modifié le fichier [11] ! Un tel système ne passait évidemment pas à l'échelle et a donc été remplacé par le DNS, passant ainsi de la classe 1 à la classe 2.

Les différents serveurs IRC, comme Freenode [12] sont également de classe 1. Sur chaque serveur, la politique technique (Transport Layer Security (TLS) ou pas, par exemple) et l'authentification (obligatoire ou pas) des utilisateurs sont communes à toutes les machines qui forment le serveur [13].

3 Classe 2

Ces systèmes répartis de classe 2 sont sans doute parmi les plus difficiles à décrire et à analyser car ils sont situés quelque part entre les extrêmes que sont les systèmes de classe 1 et ceux de classe 4. La classe 2, ce sont ceux où la modification de l'état ne dépend pas d'une organisation unique mais où il existe quand même un point [14] qui a un rôle particulier et dont la défaillance technique, ou bien des décisions politiques, peuvent affecter tout le monde.

Le DNS est un exemple d'un système de classe 2. Chaque zone est autonome, elle définit sa politique d'enregistrement de noms, et elle enregistre ou supprime des noms sans rien demander aux autres. Si je suis le titulaire de jres.org, je peux créer demain jabber.jres.org sans rien demander à personne, ni à l'Internet Corporation for Assigned Names and Numbers (ICANN), ni à Public Interest Registry (PIR).

Mais le DNS est arborescent : une zone dépend des zones situées plus haut dans l'arbre, et donc tout le monde dépend de la racine. Au moins en théorie, la racine peut supprimer un Top-Level Domain (TLD) et un registre d'un TLD peut supprimer un nom, ou le détourner [15] L'autonomie des zones est donc limitée.

Les remarques sur le DNS plus haut portaient surtout sur l'aspect politique. Et sur l'aspect technique ? Certes, chacun peut installer son résolveur DNS (ce qui est utile en cas de censure sur les résolveurs [6] [7]) mais la résolution d'un nom en données nécessite quand même un passage par la racine [16], et par les serveurs des TLD. La résolution technique DNS est donc également de classe 2.

Les systèmes arborescents sont très pratiques car ils permettent de développer des systèmes répartis, décentralisés et efficaces à la fois. Ils sont tous de classe 2 en raison de la dépendance indirecte vis-à-vis d'une ou plusieurs racines. C'est le cas par exemple des Autorité de Certification (AC) de X.509. Le système n'est pas centralisé. Une AC peut signer un certificat d'une AC secondaire, qui peut alors émettre ses propres certificats [17]. Mais elle reste dépendante de la chaîne établie depuis l'AC parente.

11. Ce système a été décrit dans [4] et [5].
12. http://www.freenode.net/
13. Le cas d'IRC est compliqué car les machines physiques qui composent le serveur peuvent, dans une certaine mesure, travailler en autonomie quand le réseau est partagé - *split*. Mais la politique reste bien unique.
14. Ou plusieurs.
15. Et cela arrive : voir par exemple l'affaire Roja Directa http://www.nextinpact.com/news/73442-rojadirecta-recupere-ses-com-et-org-illegalement-saisis-par-etats-unis.htm ou bien l'affaire Milka https://fr.wikipedia.org/wiki/Milka_contre_Kraft_Foods.
16. Notez que la gestion technique des serveurs racine est un aspect important mais très sous-documenté. Douze organisations gèrent un serveur DNS racine et, contrairement à ce qu'on lit parfois, elles ne dépendent pas de l'ICANN.
17. Parfois sous certaines contraintes définies par l'AC parente.

4 Classe 3

La classe 3 regroupe les systèmes où il faut un accord, typiquement obtenu par une forme de vote, des pairs pour participer. Aucune machine ne joue un rôle particulier, personne n'est « plus égal que les autres » mais on ne peut pas se connecter au système sans un accord des autres participants, ou d'une majorité qualifiée d'entre eux.

Largement déployé, Bitcoin est un bon exemple de système de classe 3. On ne peut pas « jouer » tout seul, les modifications qu'on apporte à la *blockchain* doivent être approuvées par les autres. À noter que Bitcoin, aujourd'hui, garde un point central : son code, dont il n'existe qu'une version. Au contraire, Ethereum [18] est bien de classe 3, car il sépare le code et la mise en œuvre. Il y a ainsi plusieurs logiciels possibles pour faire un nœud Ethereum et tous coopèrent harmonieusement.

Tous les systèmes où il y a une forme de vote pour accepter un nouveau membre sont également de classe 3. Par exemple, Tor est en train de développer un nouveau système de nommage des « services cachés », qui viendra compléter l'actuel système des clés cryptographiques en `.onion`. Ce futur système reposera sur un vote des nœuds Tor pour approuver des noms qui, contrairement à ceux de `.onion`, seront complètement choisis par les utilisateurs. Il sera donc de classe 3 alors que `.onion` est de classe 4.

5 Classe 4

Les systèmes répartis de classe 4 sont ceux où la modification de l'état ne dépend que de l'acteur de la modification, qui n'a pas à passer par un autre et peut donc « jouer » tout seul. On pourrait les appeler « réellement vraiment authentiquement pair à pair » si je n'avais pas décidé au début de ne pas utiliser les terme existants, trop et mal chargés.

Un exemple de système de classe 4 opérationnel depuis des années est le système de partage Freenet [19]. GNUnet [20] est un autre exemple de système de classe 4, mais peu déployé. GNUnet est un système très ambitieux, qui permet de remplacer complètement les protocoles Internet par un système sans centre et sans point spécifique. Il est surtout connu par son système de nommage, GNU Name System (GNS), qui peut utiliser, soit des clés cryptographiques, tirées au sort dans un espace très large [21], soit des noms locaux [22] qui n'ont une signification que par rapport à un nœud GNUnet nommé.

De même, un réseau « *mesh* » de machines connectées en Wi-Fi et se découvrant spontanément avant de router les paquets [23] est également un système de classe 4.

Contrairement à ce que prétendent certains zélateurs de la nouveauté [9], les systèmes de classe 4 ne sont pas une invention récente. L'Unix-to-Unix Copy (UUCP) d'autrefois était également un système de classe 4 très utilisé. Chaque nœud UUCP se nommait comme il voulait et échangeait avec ses voisins. Pour envoyer un message à un nœud non voisin, il fallait décider d'une route et la mettre dans le message [24]. Comme c'était peu pratique, UUCP a petit à petit évolué, en s'éloignant d'un système de classe 4. C'est ainsi que l'adressage s'est mis à utiliser des gros « hubs » d'échange comme UUnet (re-centralisation) ou des cartes donnant des listes de nœuds (cartes qui ajoutaient une dépendance nouvelle), se rapprochant plutôt d'un système de classe 3. Puis UUCP est passé au DNS et ses derniers utilisateurs dépendent donc aujourd'hui d'un système de classe 2.

18. https://www.ethereum.org/
19. https://fr.wikipedia.org/wiki/Freenet
20. https://gnunet.org/
21. Et donc quasi-uniques, sans autorité centrale, ni système de délégation arborescent.
22. Et donc non uniques.
23. Par exemple par un protocole comme Babel, cf. [8].
24. Comme dans Tor https://fr.wikipedia.org/wiki/Tor_(réseau) aujourd'hui, où le nœud de départ choisit la route…

6 Dépendances

La classification est d'autant plus difficile que les systèmes ont souvent une dépendance vers un système de classe « inférieure », qui les rend moins sexy qu'il ne pouvait sembler au prime abord. Ainsi, BitTorrent est normalement de classe 4 mais la quasi-totalité des usages passe par un moteur de recherche préalable (comme ThePirateBay) qui les fait passer dans la classe 1. De même, Extensible Messaging and Presence Protocol (XMPP) est de classe 4 (chacun installe son serveur et s'en sert comme il veut) mais est quasiment toujours utilisé via le DNS et passe donc en classe 2.

Comme beaucoup de systèmes dépendent du DNS [25], la classe 2 est donc très peuplée, alors que je n'en avais donné que peu d'exemples.

On peut également discuter du cas du routage Border Gateway Protocol (BGP) dans l'Internet. BGP lui-même est de classe 4 : tout le monde peut en faire. Toutefois, certaines règles développées au fil du temps dans la communauté des opérateurs réseau [26] sont appliquées largement et en font donc un système qui ressemble à la classe 3. Si, demain, la Resource Public Key Infrastructure (RPKI) est largement déployée et appliquée, BGP s'approchera peut-être même de la classe 2 [10].

Et l'Internet lui-même ? C'est un système réparti. À quelle classe appartient-il ? Comme c'est un système complexe, avec beaucoup de dépendances, j'auras tendance à dire qu'il est de classe 3 en théorie (pas de classe 4 : il faut l'accord des autres pour jouer) mais plutôt de classe 2 en pratique, vu sa dépendance vis-à-vis d'un certain nombre d'entités cruciales.

J'ai bien parlé de l'Internet, réseau d'interconnexion neutre, et pas du Web tel qu'il est utilisé par beaucoup d'individus, qui ne voient que « les plate-formes [27] » Si le seul usage qu'on connait est Facebook [28], on retombe dans la classe 1.

7 Conclusion

Le fait d'avoir numéroté les classes présentées plus haut peut donner l'impresssion d'un classement (avec la classe 1 la moins bonne et la 4 la meilleure ?) Mais cela serait une vision erronée. La classification est censée être objective, et n'implique pas un jugement de valeur.

Un numéro de classe supérieur ne veut donc pas forcément dire que le système est « meilleur » [29]. Ainsi, les systèmes de classe 4 ont l'avantage d'assurer à leurs utilisateurs l'indépendance maximum mais ils sont en revanche très vulnérables aux pairs qui ne « jouent pas le jeu », par exemple qui font des attaques par déni de service. En classe 4, il n'y a personne pour vous protéger [30]. En classe 3, le vote des pairs protège contre certains comportements asociaux mais laisse vulnérable à l'« attaque des 51 % [31] ».

J'espère, plus modestement, que cet article aidera à avoir des débats de meilleure qualité sur les systèmes répartis.

8 Glossaire

AC Autorité de Certification

25. Et cette dépendance est parfois discrète : certaines configurations de nœuds Bitcoin, les mineurs, notamment, utilisent des noms de domaine et flirtent donc avec la classe 2.
26. Comme le refus de préfixes IPv6 de longueur supérieure à 48 bits, voire 32 bits pour la plupart des plages de préfixe.
27. Le terme gouvernemental pour parler des GAFA.
28. http://qz.com/333313/milliions-of-facebook-users-have-no-idea-theyre-using-the-internet/
29. D'autant plus que « meilleur » n'a de sens que par rapport à des critères… qui ne sont pas consensuels.
30. Un exemple classique est l'application FireChat, sans nœud central et où les méchants peuvent donc se glisser et écouter les communications que les utilisateurs croyaient privées http://www.wired.co.uk/news/archive/2014-06/25/firechat.
31. Lorsque la majorité se met d'accord pour tricher. C'est notamment un risque pour Bitcoin. Cette attaque est un cas particulier des attaques Sybil https://en.wikipedia.org/wiki/Sybil_attack.

BGP Border Gateway Protocol. Le protocole de routage standard entre les réseaux connectés à l'Internet.

DNS Domain Name System. Système de base de données réparti, associant à un identifiant, le nom de domaine, des informations comme les adresses IP.

GNS GNU Name System. Le système de nommage de GNUnet.

ICANN Internet Corporation for Assigned Names and Numbers. Organisme états-unien chargé par le gouvernement des États-Unis de la gestion des demandes de modification de la racine DNS.

IRC Internet Relay Chat. Ancien protocole de messagerie instantanée, toujours utilisé malgré la concurrence de XMPP.

NSA National Security Agency. Service d'espionnage numérique aux États-Unis.

PIR Public Interest Registry. Registre du TLD .org.

RPKI Resource Public Key Infrastructure. Infrastructure de certificats et de signatures sur laquelle se fonde certaines solutions de sécurisation du routage.

TLD Top-Level Domain. Domaine le plus général dans un nom de domaine, situé à la fin du nom.

TLS Transport Layer Security. Protocole de sécurité, utilisant la cryptographie.

UUCP Unix-to-Unix Copy. Très ancien protocole de communication, autrefois exploité sur lignes modems et dans ses dernières années presque uniquement sur TCP/IP.

XMPP Extensible Messaging and Presence Protocol. Protocole IETF permettant notamment la messagerie instantanée.

Bibliographie

[1] Lawrence Lessig. Code is law On liberty in cyberspace, 2000. http://harvardmagazine.com/2000/01/code-is-law-html.

[2] Stéphane Bortzmeyer. Twister, un concurrent libre et pair-à-pair pour Twitter, 2014. http://www.bortzmeyer.org/twister.html.

[3] Elizabeth "Jake" Feinler. Host tables, Top level domain names and the origin of dot com, 2010. http://www.bortzmeyer.org//files/HistoryoftheTLDs.pdf.

[4] L. Peter Deutsch. RFC 606 : Host names on-line, 1973. http://www.rfc-editor.org/rfc/rfc606.txt.

[5] J. Feinler M. Kudlick. RFC 627 : ASCII text file of hostnames, 1974. http://www.rfc-editor.org/rfc/rfc627.txt.

[6] Stéphane Bortzmeyer. Censure administrative du Web en France, un premier regard technique, 2015. http://www.bortzmeyer.org/censure-francaise.html.

[7] Conseil Scientifique de l'AFNIC. Conséquences du filtrage Internet par le DNS, 2013. http://www.afnic.fr/medias/documents/conseilscientifique/CS-consequences-du-filtrage-internet-par-le-DNS.pdf.

[8] J. Chroboczek. RFC 6126 : The Babel routing protocol, 2011. http://www.rfc-editor.org/rfc/rfc6126.txt.

[9] Stéphane Bortzmeyer. Exposé sur les réseaux sociaux « alternatifs » à Pas Sage en Seine, 2013. http://www.bortzmeyer.org/pas-sage-en-seine-rezosocios.html.

[10] B. M. Mueller et B. Kuerbis. Building a new governance hierarchy : Rpki and the future of internet routing and addressing. retrieved from internet governance project, 2010. http://internetgovernance.org/pdf/RPKI-VilniusIGPfinal.pdf.