Proposition de thèse CIFRE
par Snowpack-Sorbonne Université

Encadrant Industriel : Baptiste Polvé (Snowpack)
E-mail : baptiste.polve@snowpack.eu
Page Web : https://snowpack.eu
Encadrant académique : Sébastien Tixeuil (Sorbonne Université, CNRS, LIP6)
E-mail : Sebastien.Tixeuil@lip6.fr
Page web : http://www-npa.lip6.fr/~tixeuil

19 août 2021

Proposition de titre : Évaluation et contrôle des niveaux de sécurité et d'anonymat d'un réseau d'anonymisation et de sécurité.

1 Contexte/Vision

L'augmentation des activités en ligne au cours des dernières années a conduit à une préoccupation croissante concernant la vie privée et l'anonymat des personnes. En conséquence, les solutions d'anonymat telles que Tor [8] ("The Onion Router"), I2P [13] ("Invisible Internet Project"), JonDonym [2] ou Nym [7] sont couramment utilisés par les utilisateurs d'Internet pour assurer, dans une certaine mesure, la confidentialité. Plus précisément, on cherche à masquer la source, la destination et la nature de la communication. Du point de vue des utilisateurs, ces outils permettent de naviguer sur le Web ou d'exécuter des applications en ligne sans révéler leur identité ou la teneur de leurs échanges à des tiers observant le trafic réseau. Si ces solutions reposent sur des architectures très différentes, l'essentiel des garanties de sécurité qu'elles fournissent reposent sur des primitives cryptographiques réputées inviolables.

Le réseau commercialisé par Snowpack [11] propose une approche originale qui permet d'isoler l'ensemble des éléments structurants d'une communication (Expéditeur, Destinataire, Données) pour ne laisser à un attaquant la possibilité d'attaquer, que si il parvient à contrôler l'ensemble des noeuds du réseaux ; ce qui en pratique est rendu impossible avec les technologies actuelles. A la différence des solutions précédentes, les garanties d'anonymat et de confidentialité fournies par le réseau Snowpack ne sont pas seulement basées sur des primitives cryptographiques : même si l'architecture cryptographique est compromise, l'essentiel des garanties persiste.

Malgré le nombre important de métriques relatives aux problématiques d'anonymat développées à ce jour dont près d'une dizaine sur les questions réseau [12], celles-ci tendent à adresser une architecture particulière et ne permettent pas une comparaison réelle et objective en terme d'anonymat de l'ensemble des solutions précitées les unes par rapport aux autres. Or, les niveaux de sécurité et d'anonymat dépendent de paramètres systémiques(nombre de noeuds, d'utilisateurs, volume et répartition du trafic, de liens entre les noeuds, qualité de ces noeuds) relatifs à l'architecture du réseau, mais aussi aux propriétés du protocole et des différents types de noeuds. La connaissance et la possibilité d'évaluer, notamment en temps réel, ce type de métriques est stratégique pour Snowpack dans le but d'offrir la capacité à ses utilisateurs de configurer le compromis sécurité performance le plus optimal par rapport à leurs besoins.

Pour cela, il apparaît nécessaire d'explorer la possibilité de bâtir une méthodologie unifiée permettant d'une part de comparer entre elles les solutions d'anonymat actuelles et futures, mais aussi les solutions visant à protéger la teneur des échanges, qu'elles soient basées sur des techniques de chiffrement, des communications quantiques, ou des solutions plus dynamiques telles que celle fournie par Snowpack.

2 Objectifs/Approches

L'objectif est de permettre à Snowpack de comparer tant d'un point de vue quantitatif que qualitatif son réseau d'anonymisation et de confidentialité de manière objective face à la concurrence tant des solutions d'anonymat que des solutions de confidentialité des échanges existantes.

2.1 Etat de l'art et standardisation des métriques

Le travail commencera par un état de l'art des métriques permettant de caractériser qualitativement les garanties offertes par les réseaux d'anonymisation et de confidentialité des échanges, pour en extraire les plus pertinentes ou en développer de nouvelles. Ces métriques devront permettre de comparer les différents réseaux d'anonymisation correctement établis et reconnus ainsi que le réseau commercialisé par Snowpack selon des critères communs et objectifs. Au final, une méthodologie détaillée de comparaison multi-niveau (point de vue utilisateur, point de vue prestataire de service, point de vue opérateur du réseau) sera produite.

Une fois cette méthodologie établie, une comparaison détaillée des solutions existantes sera produite sur tous les critères identifiés comme pertinents. Cette comparaison sur des critères qualitatifs permettra de positionner les solutions de l'état de l'art selon les propriétés d'anonymat et de confidentialité qui persistent en présence d'adversaires de puissance croissante.

2.2 Modélisation du réseau Snowpack et preuve mathématique des garanties offertes

Les réseaux populaires d'anonymat (Tor, I2P, JonDonym) ont fait l'objet d'études académiques qui ont abstrait les algorithmes sous-jacents, et prouvé les garanties offertes (ou leur absence) vis à vis d'adversaires clairement identifiés et mathématiquement modélisés. Ce travail n'a pas été fait pour le réseau proposé par Snowpack. Un objectif théorique important est donc de modéliser le réseau Snowpack sous la forme d'un algorithme distribué, et de considérer des adversaires de puissance croissante pour comprendre précisément quelles sont les garanties offertes, et apporter une preuve mathématique que ces garanties sont vraies pour un adversaire donné. L'expertise de Sorbonne Université sera déterminant pour mener à bien cet objectif, l'encadrant académique ayant une expérience internationale dans l'analyse d'algorithmes distribués en présence d'adversaires malveillants sans utiliser de primitives cryptographiques [3, 4, 5, 6].

A l'issue de cette modélisation, il sera possible de positionner la solution proposée par Snowpack par rapport à l'état de l'art suivant plusieurs métriques, et d'identifier les points essentiels à surveiller lors d'une implémentation.

2.3 Comparaison quantitative des solutions d'anonymat et de protection des échanges

En complément des études théoriques menées en 2.1 et 2.2, il conviendra de mettre en place une infrastructure de simulation permettant l'évaluation *quantitative* des solutions d'anonymat et de confidentialité.

Cette infrastructure de simulation devra être suffisamment générale pour évaluer les quatre solutions principales (Tor, I2P, VPN, et Snowpack) selon des critères communs : débit, latence, fraction des communications réalisées anonymement, fraction des communications réalisées confidentiellement, quantité d'information échangée sur le réseau, etc.

Une caractéristique originale de l'infrastructure de simulation sera d'injecter des attaques correspondant à un modèle d'adversaire. Plusieurs adversaires sont ainsi envisagés pour l'ensemble des solutions à tester : écoute aléatoire d'un nœud du réseau, écoute d'un nœud particulier du réseau (connaissant la topologie), écoute aléatoire de k nœuds, écoute de k nœuds particuliers du réseau (connaissant la topologie), etc. Il conviendra d'évaluer dans chaque cas si l'adversaire est capable de retrouver, pour chaque communication, l'identité de l'émetteur, l'identité du récepteur, le contenu (possiblement partiel) des communications, etc. Sorbonne Université a une expérience avérée dans la conception d'environnements de test en présence d'attaquants [1, 10, 9].

Cette évaluation quantitative des protocoles devrait permettre de mieux les paramétrer (par exemple, sur le nombre de chemins disjoints utilisés) pour résister à un adversaire particulier, et de vérifier dans quelle mesure la résistance est avérée.

3 Organisation du travail de thèse

Cette thèse est proposée dans la startup Snowpack et impliquera des interactions avec tous ses chercheurs. Au LIP6, des interactions avec les chercheurs spécialisés dans l'analyse de la tolérance aux attaques dans les réseaux et systèmes distribués, ainsi que des chercheurs spécialistes des plateformes d'expérimentation seront privilégiées. Le planning initial de cette thèse peut être divisé en quatre tâches principales :

1. Etat de l'art initial $[T0 \rightarrow T0 + 9]$. Analyse des publications en rapport avec les objectifs de la thèse, élaboration d'une méthodologie de comparaison des solutions d'anonymat et de protection des échanges.
2. Modélisation et preuve de la solution Snowpack $[T0 + 4 \rightarrow T0 + 22]$. Modélisation algorithmique, et preuve mathématiques des garanties offertes en présence de divers adversaires standard de la littérature.
3. Analyse quantitative des solutions d'anonymat et de protection des échanges $[T0 + 10 \rightarrow T0 + 31]$. Mise en place d'un environnement expérimental de test de solutions d'anonymat et de protection des échanges. Campagne d'expérimentation. Détermination des paramètres des solutions pour optimiser les garanties et/ou l'efficacité.
4. Rédaction du manuscrit de thèse $[T0 + 31 \rightarrow T0 + 36]$. Rédaction du manuscrit de thèse et préparation de la soutenance.

Chacune des tâches donnera l'opportunité de publier dans des conférences internationales ou des journaux scientifiques. Cette thèse étant proposée dans un milieu industriel, nous attendons aussi plusieurs propositions de brevets sur ces activités. Les résultats obtenus par la thèse seront intégrés aux produits Snowpack par l'équipe de R&D de l'entreprise, notamment sur les aspects représentation des niveaux de sécurité et d'anonymat en temps réel et contrôle des métriques. Ainsi, une approche permettant la mise en valeur objective de la technologie sera suivie.

Références

[1] Ali Asim and Sébastien Tixeuil. Advanced faults patterns for WSN dependability benchmarking. In Violet R. Syrotiuk, Fatih Alagöz, Brahim Bensaou, and Özgür B. Akan, editors, *Proceedings of the 13th International Symposium on Modeling Analysis and Simulation of Wireless and Mobile Systems, MSWiM 2010, Bodrum, Turkey, October 17-21, 2010*, pages 39–48. ACM, 2010.

[2] O. Berthold, H. Federrath, and M K ohntopp. Project "anonymity and unobservability in the internet". In *Proc. Workshop on Freedom and Privacy by Design / Conference on Freedom and Privacy 2000*, page 57–65, Toronto, Canada, April 2000. Association for Computing Machinery.

[3] Silvia Bonomi, Jérémie Decouchant, Giovanni Farina, Vincent Rahli, and Sébastien Tixeuil. Practical byzantine reliable broadcast on partially connected networks. *CoRR*, abs/2104.03673, 2021.

[4] Silvia Bonomi, Giovanni Farina, and Sébastien Tixeuil. Multi-hop byzantine reliable broadcast with honest dealer made practical. *J. Braz. Comput. Soc.*, 25(1) :9 :1–9 :23, 2019.

[5] Silvia Bonomi, Giovanni Farina, and Sébastien Tixeuil. Boosting the efficiency of byzantine-tolerant reliable communication. In Stéphane Devismes and Neeraj Mittal, editors, *Stabilization, Safety, and Security of Distributed Systems - 22nd International Symposium, SSS 2020, Austin, TX, USA, November 18-21, 2020, Proceedings*, volume 12514 of *Lecture Notes in Computer Science*, pages 29–44. Springer, 2020.

[6] Silvia Bonomi, Giovanni Farina, and Sébastien Tixeuil. Broadcasting information in multi-hop networks prone to mobile byzantine faults. In Chryssis Georgiou and Rupak Majumdar, editors, *Networked Systems - 8th International Conference, NETYS 2020, Marrakech, Morocco, June 3-5, 2020, Proceedings*, volume 12129 of *Lecture Notes in Computer Science*, pages 112–128. Springer, 2020.

[7] Claudia Diaz, Harry Halpin, and Kiayias Aggelos. The nym network the next generation of privacy infrastructure. February 2021.

[8] Roger Dingledine, Nick Mathewson, and Paul F. Syverson. Tor : The second-generation onion router. In Matt Blaze, editor, *Proceedings of the 13th USENIX Security Symposium, August 9-13, 2004, San Diego, CA, USA*, pages 303–320. USENIX, 2004.

[9] William Hoarau, Pierre Lemarinier, Thomas Hérault, Eric Rodriguez, Sébastien Tixeuil, and Franck Cappello. FAIL-MPI : how fault-tolerant is fault-tolerant mpi ? In *Proceedings of the 2006 IEEE International Conference on Cluster Computing, September 25-28, 2006, Barcelona, Spain*. IEEE Computer Society, 2006.

[10] William Hoarau, Sébastien Tixeuil, and Fabien Vauchelles. FAIL-FCI : versatile fault injection. *Future Gener. Comput. Syst.*, 23(7) :913–919, 2007.

[11] Frédéric Laurent and Alexis Olivereau. Dispositif et procede de transmission de donnees, April 2018.

[12] Isabel Wagner and David Eckhoff. Technical privacy metrics : A systematic survey. *ACM Comput. Surv.*, 51(3) :57 :1–57 :38, 2018.

[13] Bassam Zantout and Ramzi Haraty. I2p data communication system. In *Proceedings of ICN 2011, The Tenth International Conference on Networks*, January 2011.