

[COVID Information Commons \(CIC\) Research Lightning Talk](#)

[Transcript of a Presentation by Jaideep Vaidya \(Rutgers University-Newark\), April 14, 2021](#)



[Title: RAPID: Privacy-Preserving Crowdsensing of COVID-19 and its Sociological and Epidemiological Implications](#)

[Jaideep Vaidya CIC Database Profile](#)

[NSF Award #: 2027789](#)

[YouTube Recording with Slides](#)

[April 2021 CIC Webinar Information](#)

[Transcript Editor: Julie Meunier](#)

---

Transcript

*Slide 1:*

Bonjour à tous. Je suis Jaideep Vaidya. Je suis le directeur de l'Institut de science des données, d'apprentissage et d'applications de Rutgers, et c'est essentiellement notre réponse à la pandémie, vous savez ce que l'institut a essayé de faire pour aider à cet égard. Et c'est un travail conjoint avec mes collègues de l'École de santé publique et de l'École de communication et d'information de Rutgers.

*Slide 2:*

Donc, je voulais parler de l'un des problèmes clés que nous avons constatés tout au long de l'année dernière alors que la pandémie se déroulait, et c'était le problème des tests insuffisants et de la préparation insuffisante. Et efficacement, cela était visible partout, vous savez, cela a été rapporté par tous les médias, qu'ils soient conservateurs ou libéraux. En fait, le New York Times a publié cela : votre État fait-il suffisamment de tests de coronavirus et la réponse, bien sûr, était non au début de la pandémie, mais pendant une longue période, tout au long du sommet également, et même le Wall Street Journal a signalé que nous devons vraiment effectuer une quantité suffisante de tests COVID pour relancer l'économie.

*Slide 3:*

Donc, d'un point de vue technologique, la technologie a essayé de venir en aide à cet égard, compte tenu du manque de ressources. Et ce qui s'est passé, c'est que nous avons vu pas mal d'applications comme "COVID Near You" ou "How We Feel" qui sont apparues, qui étaient essentiellement des applications de suivi des symptômes où vous signaleriez vos symptômes, et ces informations seraient agrégées et rapportées. Même l'application CMU, que Facebook a lancée, avait cette carte interactive et ce tableau de bord. Vous mettiez vos informations, elles étaient ensuite agrégées, puis publiées, essentiellement.

*Slide 4:*

Un énorme problème que nous avons constaté à cet égard était la vie privée. Essentiellement, en fournissant ces informations, il y avait beaucoup de préoccupations, et à juste titre, concernant ce qui arriverait à votre vie privée lorsque vous signaleriez de telles informations. Et en fait, il y avait un deuxième problème aussi : les applications définissaient essentiellement des régions prédéfinies. Vous receviez des rapports pour votre ville, pour votre État, pour le comté, mais à des niveaux fixes.

*Slide 5:*

Ce que nous voulions examiner, c'était deux choses différentes. Tout d'abord, répondre à cette question sur des zones spécifiques. Si vous vous intéressez à votre quartier local ou à une région particulière d'intérêt, que se passait-il avec les symptômes ? Donc, supposez que vous ayez cette question : combien de personnes ont signalé des symptômes au cours des huit derniers jours à New York ? Ou vous vouliez regarder votre quartier local ou tout l'État, etc. Nous voulions un nombre arbitraire de requêtes portant sur les tailles et les régions. Je pourrais demander ceci, je pourrais demander cela, peut-être toute la carte là-bas, quelque chose de petit là-bas, ou que sais-je. Mais vous réalisez rapidement que dès que vous permettez ce genre de requêtes, vous savez, n'importe quel nombre de telles requêtes, le problème de la vie privée est en réalité beaucoup plus amplifié également. Essentiellement, aucun mécanisme simple d'agrégation et de publication ne pourra jamais protéger pleinement votre vie privée. En fait, avec un très petit nombre de requêtes, je peux rapidement savoir si, vous savez, des informations particulières vous concernant ou si vous contribuez ou non à vos données à une enquête de ce type.

*Slide 6:*

Alors, ce que nous avons mis au point, c'est en fait une application et un cadre appelés "COVID Nearby", et essentiellement, ce que nous fournissons, c'est une garantie formelle appelée vie privée différentielle. C'est un modèle de pointe que le recensement américain utilise - en fait, les entreprises l'utilisent également - et il garantit essentiellement, je ne vais pas entrer dans les détails, mais intuitivement parlant, il garantit que que vous mettiez ou non vos informations dans une telle application ou une telle collecte de données, toute enquête, votre risque, ou le risque pour votre identité, n'est pas amplifié. Vous n'avez pas un risque plus grand simplement parce que vous faites partie de l'enquête. Cela va être essentiellement très similaire, même si vous ne participez pas à cela. Donc, ce que nous voulions utiliser, c'était la vie privée différentielle, et c'est une garantie formelle, c'est très bien car cela garantit votre vie privée contre les attaques de confidentialité aujourd'hui, mais aussi les attaques de confidentialité qui peuvent être inventées bien à l'avance dans le futur dont nous ne savons rien. Un problème, cependant, est que dès que vous essayez de le faire, surtout sur de telles données spatio-temporelles et/ou dynamiques, il est très difficile de fournir une utilité, de bonnes réponses, tout en garantissant la vie privée, même si vous faites de la vie privée différentielle. Donc, ce que nous avons fait, c'est que nous avons trouvé un moyen unique de représenter les données. Nous avons créé cette application qui vous permettra de signaler vos symptômes et qui sera stockée dans une base de données sécurisée, nous agissons en tant que conservateur de confiance ici, mais il y a une représentation intermédiaire qui est construite sur ces données, qui est ensuite utilisée pour répondre à un nombre quelconque de questions sur les données pour garantir que, peu importe, vous savez, même si vous posez mille questions, le risque de confidentialité pour quiconque ne cesse pas d'augmenter. Il est essentiellement limité là-bas.

*Slide 7:*

Et cette application a effectivement été développée, il y avait un tas de défis comme je l'ai dit au niveau technique. Ce sont des données spatio-temporelles. Elles changent constamment. Vous avez un nombre arbitraire de requêtes et en fait, si vous voulez que les gens l'utilisent, il doit vraiment être capable de fonctionner rapidement également. Les résultats doivent être utilisables, et ils doivent fonctionner réellement pour la situation spécifique en cours, dans ce cas, la pandémie de COVID. Et étonnamment, vous savez, en dehors des défis techniques, l'un des plus grands défis était avec l'effort de développement rapide, l'obtenir certifié par le Play Store et l'App Store [Apple] afin que nous puissions l'avoir là-bas. Mais nous avons réussi à le faire et en fait, l'application est là-bas et si vous avez un téléphone Android ou un appareil Apple iOS, vous pouvez absolument l'obtenir et y mettre vos données en toute garantie de confidentialité. Une chose que je voulais souligner - nous avons examiné ces données et, oui, nous avons passé par les CER [Comités d'éthique de la recherche] et tout cela. La bonne nouvelle, c'est que vous pouvez en fait obtenir des résultats très proches des résultats non privés en utilisant également ces techniques. Donc, ces barres que vous voyez sont essentiellement les données originales, puis les données privées. Les barres empilées vous donnent une idée qu'elles sont à peu près les mêmes, mais juste pour vous donner une meilleure vue à une date particulière, par exemple, vous pouvez voir la carte thermique pour le privé par rapport à l'original, et comme vous pouvez le voir, fondamentalement, le classement relatif, non seulement au niveau du comté mais même à l'intérieur des comtés, s'avère être exactement le même. Donc, encore une fois, je n'ai pas le temps d'entrer dans les détails de cela, mais en termes de suivi des points chauds ou de classement, il est très facile d'obtenir des résultats exactement similaires, et vous ne perdez pas en précision en faisant cela.

*Slide 8:*

Donc, avec cela, je m'arrêterai essentiellement et dirai simplement, oui, vous devriez utiliser l'application et les liens sont là si vous voulez l'utiliser sur votre appareil Apple ou votre appareil Android, et je m'arrêterai là. Merci.