The True Random Privacy Project

True Random Privacy (TRP) project aims to create, during RaP! 2020, a new differential privacy solution for images, embedding a state-of-the-art features description technique.

Differential privacy (DP) is a property of data distributions that limits the amount of information released by a specific individual enabling the public sharing of data [1]. This is achieved through the modification of content to withhold the specificities of the individual while releasing population-wide features. However, pseudo-random generators used to generate noise have large scale patterns that can be spotted, compromising the privacy. A key component of the success of differential privacy is in the true randomness of the added noise [2] [3].

Differential privacy

During the hackathon, we will the use quantum generated numbers (true random) given by RaP! to apply a differentially private speckle noise disrupting images, creating data obfuscation, to a point where the identification of their source will become impossible. While DP has suffered criticism for its reliance on noise for its expression, and the lack of utility of the resulting data, most of it relies on the assumption of practical implementation of mechanisms with pseudorandom noise. In practical terms, True Random DP achieves better results by reducing the amount of added noise to achieve the same result, increasing the populational information redundancy, and thus permitting to use DP efficiently.

While data sharing between healthcare institutions has been proven to be essential for the development of research on rare diseases and new pandemics such as COVID-19, current cryptographic measures remain ineffective or limiting. Differential privacy tackles this problem permitting a data sharing while keeping a large part of the utility of the data. In this hackathon, we will ensure that the level of privacy and security of the mechanisms is ensured based on mathematical proofs and in close contact with data security experts.

The feature identification for comparison

The stronger the noise, the better the privacy, the more difficult the data exploitation. To make our results exploitable, we will use a new approach for features identification developed by members of our team [4]. This approach takes profit of the random distribution of the noise applied to detect large features and describe them. The features of privatized images can then be then compared, classified, clustered, depending on the application needs.

This combination of quantum number generation and feature identification will allow a novel and unique product for sharing sensitive data and exploiting them!

[1] C. Dwork and A. Roth, The Algorithmic Foundations of Differential Privacy, 2014.

[2] Y. Dodis, López-Alt, Adriana, Mironov, Ilya and Vadhan, Salil, "Differential Privacy with Imperfect Randomness," 2012.

[3] Garfinkel, Simson L and Leclerc, Philip, "Randomness Concerns When Deploying Differential Privacy," 2020.

[4] Y. Donon, "Key points detection algorithm for noised data," CERN, Geneva, 2020.

keypoints detection and data governance. The three of us are data scientists eager to use our skills to make the world a little better.

We see in the Random Power Hackathon a great opportunity to bring our knowledge together. We want to create an application presenting both scientific novelties and being useful for the industry, out of the box, using random generation techniques of tomorrow.

José Cabrero-Holgueras, 24 years old and coming from Spain, is a Ph.D. candidate at CERN Openlab and University Carlos III of Madrid. He works on the improvement of Privacy Enhancing Techniques for its usage in Machine and Deep Learning enabling private training and prediction of statistical models. The focus of his research is in the healthcare sector, where he seeks to enhance personalized medicine through the introduction of collaborative techniques between different healthcare institutions.

Dr. Rustam Paringer, 30 years old and joining us from Russia, is a senior researcher at the Technical Cybernetics department in Samara National Research University (Russia) and at the Image Processing Systems Institute of the Russian Academy of Sciences - branch of the FSRC RAS. His main research interests include computer image processing, pattern recognition and data mining.

Dr. Yann Donon, 27 years old and from Switzerland, recently earned a joint Ph.D. between Samara National Research University and CERN. He works in both institutions and at Image Processing Systems Institute of the Russian Academy of Sciences – branch of the FSRC RAS as a researcher. Yann is also IT security Officer in the Swiss armed forces and worked four years in the medical industry, two places where he grew a passion for data privacy and security. His research interest includes behaviour prediction, noised data manipulation and data governance.

Team contact:
Yann Donon
yann.donon@cern.ch
+41798706289