PRESS OFFICE • 1 MetroTech Center, 19th Floor, Brooklyn, NY 11201

CONTACT • Karl Greenberg
646.997.3802 / mobile 646.519.1996
Karl.Greenberg@nyu.edu

**Immediate Release**

# Team streamlines neural networks to be more adept at computing on encrypted data

BROOKLYN, New York, Wednesday, June 23, 2021 – Researchers at the NYU Center for Cyber Security at the NYU Tandon School of Engineering are rethinking basic functions that drive the ability of neural networks to make inferences on encrypted data. Their focus is on linear and non-linear operators, key features of neural network frameworks that, depending on the operation, introduce a heavy toll in time and computational resources. When neural networks compute on encrypted data, many of these costs are incurred by rectified linear activation function (ReLU), a non-linear operation.

Brandon Reagen, professor of computer science and engineering and electrical and computer engineering and a team of collaborators including Nandan Kumar Jha, a Ph.D. student, and Zahra Ghodsi, a former doctoral student under the guidance of Siddharth Garg, developed a framework called DeepReDuce that offers a solution through rearrangement and reduction of ReLUs in neural networks. The investigators will present their paper "DeepReDuce: ReLU Reduction for Fast Private Inference" at the 38th International Conference on Machine Learning (ICML 21) in July, 2021.

Reagen explained that this shift requires a fundamental reassessment of where and how many components are distributed in neural networks systems.

"What we are trying to do is rethink how neural nets are designed in the first place," he explained. "You can skip a lot of these time- and computationally-expensive ReLU operations and still get high performing networks at 2 to 4 times faster run time."

The team found that, compared to the state-of-the-art for private inference, DeepReDuce improved accuracy and reduced ReLU count by up to 3.5% and 3.5×, respectively.

The inquiry is not merely academic. As the use of AI grows in concert with concerns about the security of personal, corporate, and government data security, neural networks are increasingly making computations on encrypted data. In such scenarios involving neural networks generating private inferences (PI's) on hidden data without disclosing inputs, it is the *non-linear* functions that exert the

highest “cost” in time and power. Because these costs increase the difficulty and time it takes for learning machines to do PI, researchers have struggled to lighten the load ReLUs exert on such computations.

The team’s work builds on innovative technology called CryptoNAS**.** Described in an earlier paper whose authors include Ghodsi and a third Ph.D. student, Akshaj Veldanda, CryptoNAS optimize**s** the use of ReLUs as one might rearrange how rocks are arranged in a stream to optimize the flow of water: it rebalances the distribution of ReLUS in the network and removes redundant ReLUs.

DeepReDuce expands on CryptoNAS by streamlining the process further. It comprises a set of optimizations for the judicious removal of ReLUs after CryptoNAS reorganization functions. The researchers tested DeepReDuce by using it to remove ReLUs from classic networks, finding that they were able to significantly reduce inference latency while maintaining high accuracy.

Reagan, with Mihalis Maniatakos, research assistant professor of electrical and computer engineering, is also part of a collaboration with data security company Duality to design a new microchip designed to handle computation on fully encrypted data.

The research on ReLUS was supported by ADA and the Data Protection in Virtual Environments (DPRIVE) program at the U.S. Defense Advanced Research Projects Agency (DARPA) and the Center for Applications Driving Architectures.

.

***About the New York University Tandon School of Engineering***

*The NYU Tandon School of Engineering dates to 1854, the founding date for both the New York University School of Civil Engineering and Architecture and the Brooklyn Collegiate and Polytechnic Institute. A January 2014 merger created a comprehensive school of education and research in engineering and applied sciences as part of a global university, with close connections to engineering programs at NYU Abu Dhabi and NYU Shanghai. NYU Tandon is rooted in a vibrant tradition of entrepreneurship, intellectual curiosity, and innovative solutions to humanity’s most pressing global challenges. Research at Tandon focuses on vital intersections between communications/IT, cybersecurity, and data science/AI/robotics systems and tools and critical areas of society that they influence, including emerging media, health, sustainability, and urban living. We believe diversity is integral to excellence, and are creating a vibrant, inclusive, and equitable environment for all of our students, faculty and staff. For more information, visit engineering.nyu.edu.*

###

 www.facebook.com/nyutandon

 @NYUTandon