MIT's Quantum Locks Strengthen Cloud AI Security

Posted by Okachinepa 09/30/2024

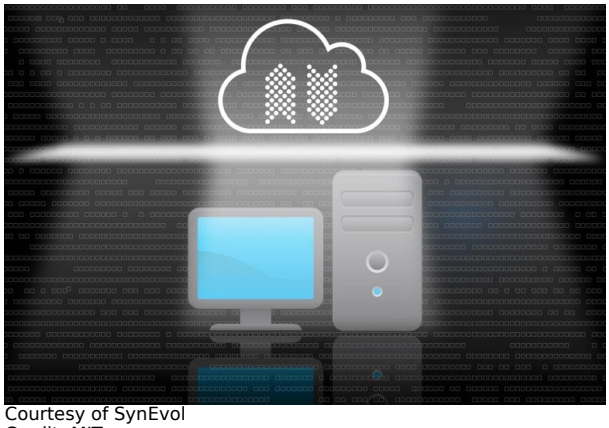
Courtesy of SynEvol
Credit:MIT

Numerous industries, including financial forecasts and healthcare diagnostics, have found use for deep learning models. Owing to their heavy processing load, these models rely on reliable cloud servers.

But this heavy reliance on cloud computing comes with a lot of security hazards. This is particularly important in delicate industries like healthcare, where hospitals may be discouraged from using AI tools due to privacy concerns around patient data.

In order to address this urgent problem, researchers at MIT have created a security protocol that makes use of light's quantum characteristics to ensure that data is sent securely to and from cloud servers during deep learning calculations.

The protocol takes advantage of the basic ideas of quantum physics to encode data into the laser light used in fiber optic communications systems. This prevents hackers from secretly copying or intercepting the data.



Courtesy of SynEvol
Credit: MIT

Furthermore, the method ensures confidentiality without sacrificing the deep-learning models' accuracy. The researcher conducted tests to show that their procedure could guarantee strong security protections and retain 96 percent accuracy.

"GPT-4 and other deep learning models offer never-before-seen capabilities, but they also demand a lot of processing power. “Our protocol allows users to take advantage of these potent models without jeopardizing the confidentiality of their data or the proprietary nature of the models themselves,” claims Kfir Sulimany, the main author of a paper on this security protocol and an MIT postdoc at the Research Laboratory for Electronics (RLE).

Prahlad Iyengar, a graduate student studying electrical engineering and computer science (EECS), Ryan Hamerly, a former postdoc at NTT Research, Inc., and senior author Dirk Englund, a professor of EECS and principal investigator of the Quantum Photonics and Artificial Intelligence Group and of RLE, join Sulimany on the paper. Recent presentations of the study were made at the Annual Conference on Quantum Cryptography.

The researchers concentrated on a cloud-based compute situation where two parties are involved: a central server that manages a deep learning model and a client that possesses sensitive data, such as medical photos.

Without disclosing any personal information about the patient, the customer wants to utilize the deep-learning model to predict things like if a patient has cancer based on medical pictures.

To make a prediction in this case, sensitive data must be sent. Nonetheless, patient data security must be maintained throughout the procedure.

Furthermore, the server is unwilling to divulge any information on the unique model that OpenAI and other companies have spent years and millions of dollars developing.

Vadlamani continues, "Both parties have something they want to hide."

A malicious party could quickly replicate data transferred from the server or client while using digital computing.

Comparatively speaking, quantum information is not completely replicable. The no-cloning principle is a feature that the researchers take advantage of in their security approach.

The server uses laser light to encode the weights of a deep neural network into an optical field for the researchers' protocol.

Layers of connected nodes, or neurons, that process data are arranged in layers within a neural network, a deep learning model. The parts of the model that do the mathematical operations on each input, layer by layer, are called weights. Up until the last layer produces a forecast, the output of one layer is transferred into the subsequent layer.

The server transmits the network’s weights to the client, which implements operations to get a result based on their private data. The information is kept hidden from the server.

In addition, due of the quantum nature of light, the security protocol forbids the client from replicating the weights and only permits the client to measure a single result.

The protocol is made to cancel out the first layer as soon as the client feeds the first result into the subsequent layer, preventing the client from learning any more about the model.

The client just measures the light required to operate the deep neural network and feed the output into the following layer, as opposed to measuring all of the incoming light from the server. The leftover light is then sent back to the server by the client for security checks, according to Sulimany.

When assessing the model's output, the client inevitably introduces small mistakes because of the no-cloning theorem. The server can measure these mistakes to find out if any information was spilled when it receives the residual light from the client. Crucially, it has been demonstrated that the remaining light conceals the client data.

Optical fibers are usually used in modern telecommunications equipment because they can sustain large bandwidth across long distances. Without the need for additional gear, the researchers can encode data into light for their security protocol because this equipment already has optical lasers.

Upon testing, the researchers discovered that their method could ensure both server and client security while allowing the deep neural network to reach 96 percent accuracy.

Less than 10% of what an adversary would need to retrieve any secret information is contained in the tiny bit of model information that is disclosed when the client executes operations. Conversely, a malevolent server could only acquire roughly 1% of the data required to pilfer the client's information.

Building on years of quantum cryptography work that had also been demonstrated on that testbed, Englund says, "a few years ago, when we developed our demonstration of distributed machine learning inference between MIT's main campus and MIT Lincoln Laboratory, it dawned on me that we could do something entirely new to provide physical-layer security." Nevertheless, a number of complex theoretical obstacles needed to be resolved before it would be possible to implement distributed machine learning with privacy guarantees. This was not conceivable prior to Kfir joining our team, since Kfir was able to establish the unified framework supporting this work by having a unique understanding of both the theory and experimental components.

In the future, the researchers hope to investigate how this protocol may be used in conjunction with federated learning, a method in which several parties train a central deep-learning model using their data. Moreover, it might be applied to quantum operations as opposed to the classical operations they researched for this work, which might offer benefits in terms of precision and security.

In particular, deep learning and quantum key distribution are two domains that are not typically combined, yet this work does so in a creative and captivating way. It provides an extra degree of security to the former and makes what seems like a realistic implementation possible by utilizing techniques from the latter. Regarding maintaining privacy in distributed architectures, this may be of interest. The practical implementation of the protocol and its behavior under suboptimal experimental conditions are exciting to watch, says Eleni Diamanti, a CNRS research director at the Sorbonne University in Paris who was not involved in this work.
