Top Brazilian Bank Pilots Privacy Encryption Quantum Computers Can’t Break

Inside IBM Research ⋮ 9-12 minutes ⋮ 1/10/2020

Using a novel crypto scheme called homomorphic encryption, researchers apply machine learning on fully encrypted banking data — without having to decrypt it first.

By Katia Moskvitch

Even a quantum computer wouldn’t be able to break homomorphic encryption. [Credit: IBM Research.]

More than 5,000 publicly disclosed data breaches and billions of personal records exposed. Hacking attacks and deliberate data exfiltration by insiders leading to leaks of sensitive financial, medical and government information. That’s only in 2019, globally — despite the data often being encrypted.

But one of the top banks in Brazil wants to avoid such a disaster. For the past year, Banco Bradesco S.A., one of Brazil’s biggest banking and financial services companies, has been working with IBM Research to apply encryption techniques of the future. The idea is to secure data in a novel and totally inpenetrable way, so much so that even a quantum computer wouldn’t be able to get to it—using an type of cryptography which was first suggested by mathematicians in the 1970s called homomorphic encryption.

Having encrypted real banking data, the scientists demonstrated secure machine learning-based predictions. They presented the results this week at the IACR Real World Crypto Symposium 2020 in

New York City.

So what *is* homomorphic encryption?

The term is a mouthful, but the approach is simple.

Cryptography “underpins the trust that we have in e-commerce, blockchains and crypto-currencies,” says Mike Osborne, a crypto researcher at IBM. Today, data at rest such as backups as well as data in transit sent between organizations or simply between one’s computer and, say, an e-commerce website, are typically encrypted.

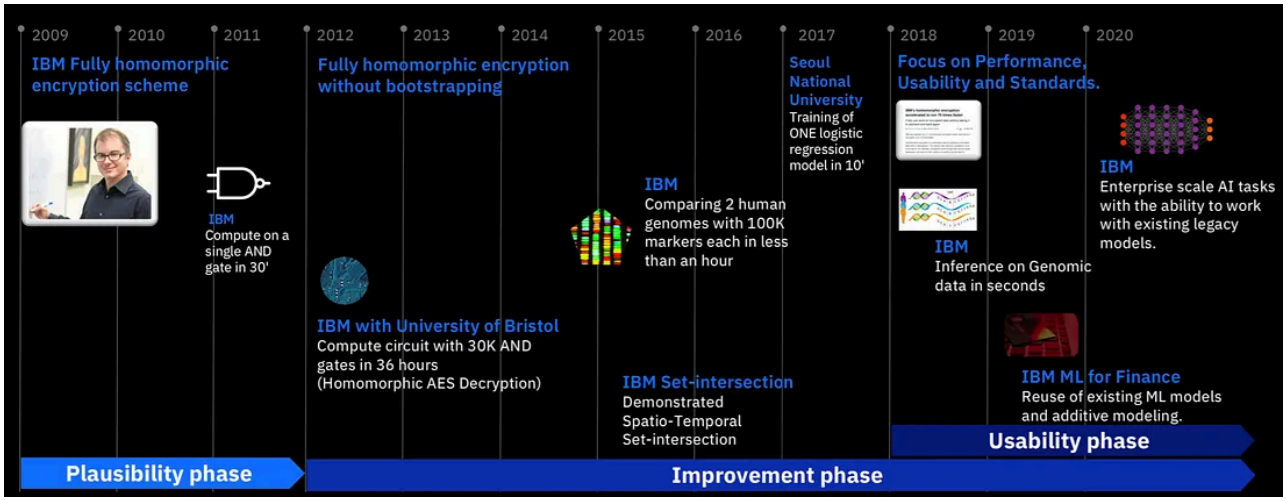
The problem is, to be processed, the data need to be decrypted first, meaning exposing the information for a certain length of time. This window, however brief, is enough for data exfiltration and leaks. That’s why before any computations, sensitive data are usually pre-processed to obfuscate, anonymize or even remove any private and sensitive fields. But for certain applications, such as in banking, removing all private and sensitive information doesn’t make sense.

But what if the people handling the data wouldn’t have to decrypt anything to perform computations? Well, then, leak or not, the information would remain safe and sound. Provided the encryption is strong enough, that is.

> IBM Research is focused on how to apply advanced cryptographic schemes such as Homomorphic Encryption (HE) to protect the privacy and confidentiality of both the data during the training of ML models as well as the models themselves, and as a consequence, the prediction task can also be protected. [Credit: IBM Research.]

That’s where homomorphic encryption comes in. Pure research since the 1970s, it’s finally crawling out of the lab and into real-world applications. If encrypted homomorphically, data stays encrypted — all the time. The information remains cryptographically jumbled to preserve privacy and confidentiality while it is being processed, so that even the people handling it inside the organization can’t know the contents.

The first fully homomorphic encryption system was developed in 2009 by IBM scientist Craig Gentry. He compared it to “one of those boxes with the gloves that are used to handle toxic chemicals… All the manipulation happens inside the box, and the chemicals are never exposed to the outside world.” As IBM researcher Flavio Bergamaschi, the lead author of the recent pilot, puts it, “this changes completely the paradigm of security and computation.”



Timeline of homomoprhic encryption. [Credit: IBM Research.]

For instance, it would be possible to use machine learning to match a patient for a clinical trial based on a their genome sequence which is homomorphically encrypted in the cloud. The system would output an encrypted a match, without exposing any sensitive data. “The query can be performed ‘obliviously,’ without revealing the data or the result to the agent processing the query,” says Dave Braines, a researcher at IBM.

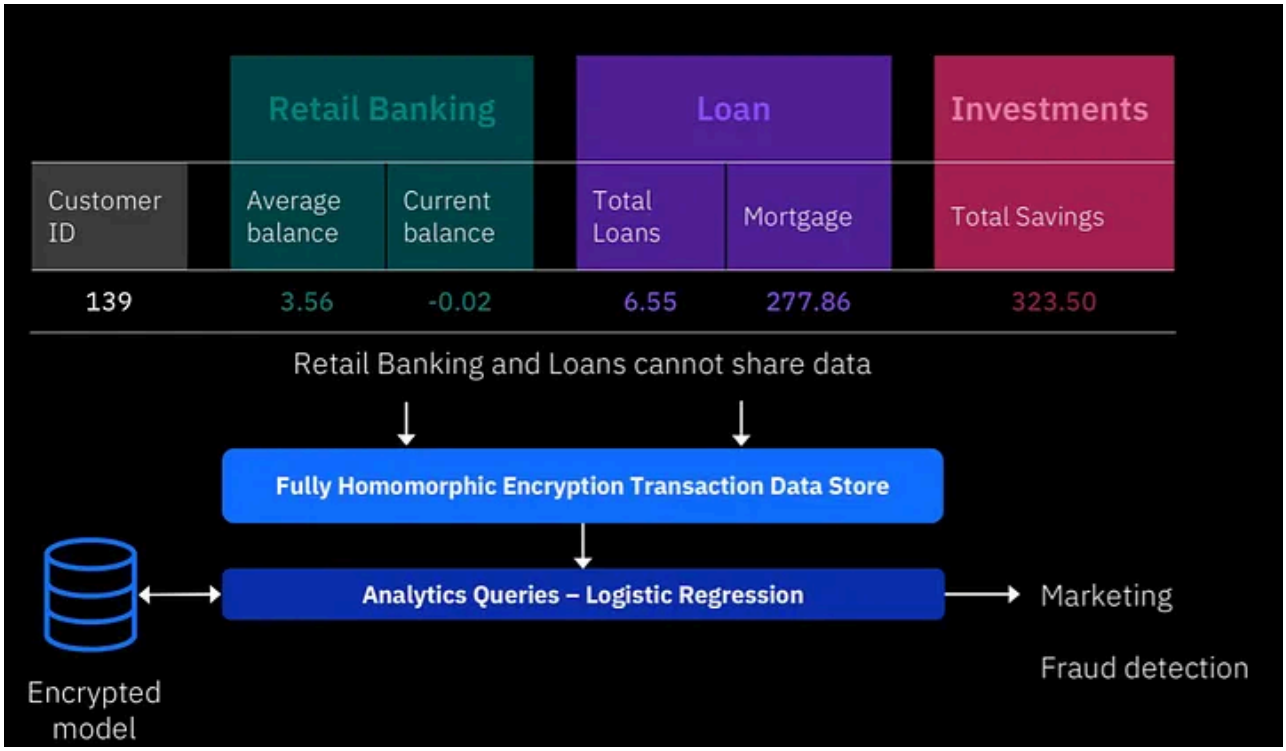
And, crucially, homomorphic encryption is resistant to any potential attacks from quantum computers. Once fully-functional, quantum computers should greatly surpass traditional computers for certain types of computations, for instance to predict financial investments and develop new drugs and materials. And they could also easily crack current encryption algorithms.

But not homomorphic ones. That's because this type of encryption is based on the mathematics of lattices — repeating, multidimensional grid-like collection of points. Lattice-based schemes hide data inside this lattice, some distance away from a point, and determining how far away an encrypted message is from a lattice point is extremely difficult for both a quantum or a traditional computer. But if you know a secret key, then it's easy to get the message.

"Developments in quantum computing have given much of the cryptography that we use today an end-of-shelf life," says Osborne. "Unfortunately, it is a common misunderstanding to think that we can wait until a large quantum computer is built before doing anything. We can safely say that homomorphic encryption is secure against quantum computers."

What's the recent pilot with Banco Bradesco?

In the latest research, the scientists worked with Banco Bradesco real financial data. They showed that it was possible to perform encrypted predictions concealing the data and the result throughout the processing, obtaining the level of privacy not currently possible with any other methods, says Bergamaschi. Homomorphic encryption "can provide privacy protection for users requesting predictions," he says— redefining the boundaries of what data must be stored and by whom.



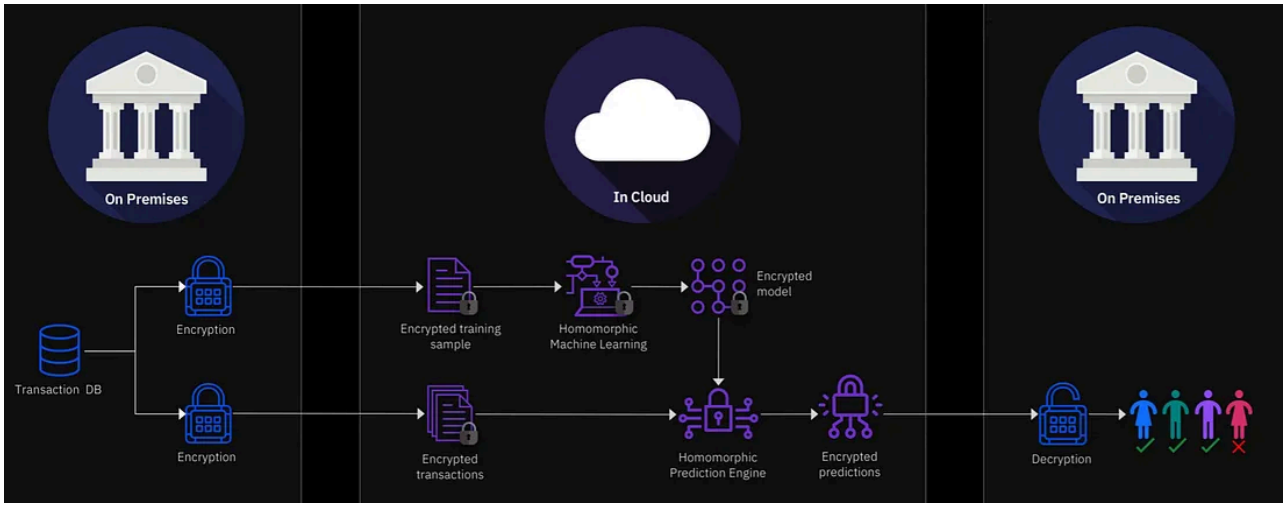
Real financial data proof-of-concept with Banco Bradesco. [Credit: IBM Research.]

The team took transaction data and an existing machine learning-based prediction model and performed two experiments. First, they homomorphically encrypted the data and the model, and showed that it was possible to run predictions with the same accuracy as without encryption. It means that banks can safely outsource the task of running predictions to an untrusted environment, adds Bergamaschi.

They then trained the model using encrypted data, showing that it was possible to use homomorphic encryption to preserve the privacy of data. Typically, financial institutions gather information on their clients — analyzing how much a person spends on groceries, petrol and so on, and predicting whether that customer would soon need a loan.

To do this, a bank's analysts usually manually identify the most important features about an individual's financial history that would allow them to make that prediction, selecting a handful of features out of about a thousand. While doing so, they have access to the data, potentially compromising it.

"We've shown that we could do this important task homomorphically with encrypted data coming in," says Bergamaschi. This way, the key features are chosen without exposing any information about the customer — meaning that there is "potential to reduce the damaging consequences of data leaks that we have seen in the past, and the breach of privacy of individuals."



Homomorphic encryption ensures secure machine learning in the cloud. [Credit: IBM Research.]

Banking is not the only application of homomorphic encryption. Whenever we search for directions on our smartphone or make a purchase online, we give away a lot of our private information by telling the service in the cloud where we want to go or what we are looking for. If we did that homomorphically, we would send the data encrypted in a way that the service wouldn't be able to 'look' inside.

> "That's the beauty of homomorphic encryption," says Bergamaschi. It would be particularly useful to safeguard an individual's medical records or genome sequence.

Homomorphic encryption also addresses the problem of sharing information. This is crucial because of regulations such as GDPR in Europe, specific privacy laws or even a company's own regulations. Again, take a bank. There are many different departments: for instance, one is dealing with insurance and another one with investment. If they were to merge, there would be data aggregation — and someone, typically the data analyst, would have access to all the data. "That's one of the weakest points for data exfiltration," says Bergamaschi — the person that has the access to all the data as part of their day job. Some banks have strict regulations about the use of external devices, but some don't — and data may get compromized.

One issue with homomorphic encryption today is that the computational requirements are way greater with encrypted data, making the process much, much longer. But the technology has been improving over the years, and now we are finally "at that point where we can achieve adequate performance when protecting the privacy and confidentiality of the data is paramount," says Bergamaschi.

IBM cryptographers have just release the Homomorphic Encryption library — HElib v1.0.0. For the past year the beta versions have been under extensive refactoring for reliability, robustness & serviceability, performance, and most importantly usability for researchers and developers working on HE and its uses. Available today in GitHub.

Towards a Homomorphic Machine Learning Big Data Pipeline for the Financial Services Sector, Oliver Masters, Hamish Hunt, Enrico Steffinlongo, Jack Crawford, Flavio Bergamaschi, Maria

Eugenia Dela Rosa, Caio Cesar Quini, Camila T. Alves, Fernanda de Souza, and Deise Goncalves Ferreira