Securing data during computation

Between US $375-575 billion is lost every year due to cybercriminal activities. **Professor Marina Blanton** is helping to improve this issue by creating systems so organisations can jointly compute on data without sharing their specifics, and by making sure that private data remain protected, even when being outsourced to the cloud

Why were you first drawn to information security, privacy and applied cryptography?

I started working on security-related topics during my Master's degree at Ohio University in the US. My research was highly empirical and allowed me to make only limited conclusions about the phenomena I was observing. As a result, for my PhD I was looking to explore areas better suited to formal analysis and rigorous scientific evaluation. My work on privacy-related topics and applied cryptography started early in my PhD studies. These topics are very important for modern society where many aspects of one's life are captured and stored in a digital form.

Can you introduce secure computation?

Secure multi-party computation refers to the ability of multiple participants to jointly compute on their respective private data without revealing information about those data to each other, except for the agreed upon computation outcome. A secure solution in this framework must guarantee that interactions are protected through provably secure cryptographic techniques, and it is not possible for any contributor or outsider to learn unintended information about other participants' data.

A range of applications benefit from such techniques; they include computation activities spanning sensitive databases that belong to different organisations (eg. medical records) and commercial applications (eg. testing for predisposition to a genetic condition without revealing one's DNA), among others.

How does this relate to secure computation outsourcing?

When such techniques allow computation to be performed securely by entities who are different from the data owners, they become suitable for secure computation outsourcing. In that setting, one or more clients utilise one or more external servers to perform computation on their private data in such a way that it is not possible for the servers to learn anything about the data they handle.

Has the advent of cloud computing increased the need for greater privacy-preserving computation?

Cloud computing is an attractive mechanism for using abundant computing and storage resources of large service providers and lowering one's operational costs. The use of external resources, however, requires clients to relinquish control of their data to third parties. Because of a lack of transparency and inability to control what happens to one's data once it is outsourced to the cloud, there is a resistance to use cloud computing for computation with proprietary, private or otherwise sensitive information. The need for privacy-preserving computation has existed for a long time, but the ubiquitous use of cloud computing has taken this need to a new level and placed unique demands on the security objectives that must be met.

Can you introduce Private dIstributed Computation COmpiler (PICCO)?

PICCO is a compiler that allows a program written in the popular programming language C to be converted into a secure implementation of the same functionality. Our goal was to enable programmers without extensive knowledge of secure computation techniques to build a solution specific to their desired computational task.

Using PICCO, a programmer specifies what variables will hold private data. Our compiler then transforms the program into an equivalent in which all private data are fully protected. The resulting program can be used for secure joint computation or secure outsourcing with provable security guarantees. The compiler also supports optimisations that help improve the runtime of the resulting secure implementation.

How does PICCO improve security of outsourced computation over conventional methods?

The current practice is to use no data protection when placing one's computation with cloud providers. Instead, clients sign a contract with the service provider that can include provisions for preventing data sharing of the client's data by the cloud provider. With PICCO, private data is never available in an unprotected form to any entity other than the data owner. Therefore, no employee at the cloud service provider will have access to clients' data. Similarly, if systems are compromised by an external attacker or malware, no one will be able to extract meaningful information about the data.

Can you provide a brief insight into your activities with biometric data protection?

Given a rapid recent increase in the use of biometric data in various applications and highly sensitive nature of such data, we develop custom solutions for securely computing with different biometric modalities or securely outsourcing such tasks. The security requirements are as strong as before, and the main goal is higher efficiency than what generic solutions provide. In addition, we develop techniques to efficiently verify that the result of outsourced computation is indeed correct.

Strong data protection

Data security is a prominent concern for businesses and governments alike, and it can prevent such organisations from making the most of the information they have to hand. Researchers at the **University of Notre Dame** are working on secure computation methods that allow outsourcing and collaboration without fear that data will be compromised

KNOWLEDGE IS POWER. Data on everyday citizens, consumers or patients has always been valuable, but in a digital world such records can be stored in quantities that would previously have been impossible. The rise of big data is a double-edged sword, in the sense that vast repositories of information represent a vulnerable target for malicious agents; indeed, there have been several high-profile cases of data theft in recent years.

And there is another challenge, too. It is very difficult to effectively mobilise data that are retained in silos. Take medical data as an example; patient records are commonly kept in disparate systems maintained by different organisations, effectively preventing them from being examined comprehensively. To compound the problem, even if collaborating organisations were permitted and willing to share their information for joint computation, their information would be more vulnerable to attack during this process.

SECURE IN THE KNOWLEDGE

The solution, according to Professor Marina Blanton and her colleagues at the University of Notre Dame, is secure computation – a processing method whereby participants can collaborate on computation without revealing their own private data. For example, imagine hospitals A, B and C each have patients with the same rare medical condition. Each on their own does not have enough data to determine the most effective treatment for the condition, but they would have enough data if they combined sources. However, because of privacy issues, legally they cannot freely share their patients' records. Blanton's work with secure computation would change this by enabling hospitals to participate in a joint computation through a secure system and share results without uncovering any private, medical details about the patients.

Secure computation must be absolutely watertight in order to be useful, because end-users must be assured that nothing about their data can be discovered. Blanton and her colleagues have been pursuing a project that could make this sought-after ability a reality – unlocking much of the value that is currently locked away in data silos all over the world. The project has led to the development of the Private dIstributed Computation COmpiler (PICCO), which is able to translate a general-purpose program, written in an extension of the C programming language, into a secure implementation. "As a community, we've made large leaps toward improving performance of secure computation and outsourcing techniques and developing necessary components to make general-purpose secure computation possible," Blanton summates.

KEEPING DATA IN THE DARK

There are many aspects to this challenging work that require the researchers' attention, but one factor that is particularly key is the development of what Blanton and her team refer to as 'data-oblivious' algorithms. With standard algorithms, many steps within the computational task depend on the original input – but this is not a secure way of handling that original input, because it allows the opportunity for data to be leaked at the processing stage. Data-oblivious execution bucks this trend by making computation steps independent of the data; the data can remain hidden with no loss of function.

The vast majority of data structures and algorithms are currently 'non-oblivious' and naive attempts to address this often result

to only be encrypted while at rest; they have to be decrypted to be used in computation, meaning they can be retrieved from the computer's memory by someone with access to that system, or they can be altered. Therefore, if data owners want to use the result of an outsourced computation, they must then verify that the computations were performed as prescribed.

SECURE OUTSOURCING AND VERIFICATION OF RESULTS

Blanton and her colleagues are working on techniques to make sure that data are secure no matter what environment they are used in. One area in which they are applying these techniques is biometric data. Biometric research often involves running computation on large amounts of data, making cloud computing very appealing in order to reduce the burden on in-house memory and processing power. However, biometric data is extremely personal data; as such, medical organisations have not been able to seize on the benefits that cloud-computing provides in the interest of maintaining privacy. "The US National Institutes of Health (NIH) used to maintain a database of anonymised DNA sequences for researchers to use, but the database was taken down after it had been shown that the individuals whose DNAs were included in the database could be re-identified," Blanton shares.

Excitingly, her work in secure computation combined with her activities in secure outsourcing could bring the database's benefits back within reach and without patient re-identification concerns; her techniques would allow collaborative and outsourced computation on data from multiple sources where no information is revealed except the final outcome. Moreover, her techniques are robust enough to ensure that the results of the outsourced computations are verifiably correct.

A COLLABORATIVE ENDEAVOUR

Blanton's secure computation techniques are not only revolutionary, but they are enabling collaboration and computation that is not possible otherwise – one only has to look at her success in developing specific methods for iris images, voice recordings, fingerprints and DNA sequences to find this to be true. Even just considering fingerprints, Blanton has created protocols that allow for secure comparison of two fingerprints from different data owners in just a few milliseconds.

Blanton is excited to continue her working with colleagues to improve secure computation, secure outsourcing and data verifiability, as well as preserving data privacy in general. "We are at the point where secure computation and secure outsourcing techniques are efficient enough for complex and diverse computations, and I expect their adoption in practice to only increase," she concludes.

in significant increases in computation time. The task for Blanton and her colleagues is to construct data-oblivious alternatives that are as close in time efficiency as possible to their traditional counterparts.

The researchers have already made significant progress in producing data-oblivious algorithms suitable for secure floating-point arithmetic and working with graphs; these algorithms have been integrated into the PICCO compiler. "The broader impact of this work includes societal benefits such as safer practices in handling private or sensitive data when it is being used in any type of outsourced and/or joint multi-party computation," Blanton enthuses. "This is directly applicable and useful to government, health, military and commercial sectors, to name a few."

THE DEAL WITH OUTSOURCED DATA

Blanton is also applying her work with secure computation to data when they are outsourced to the cloud or a storage service. "Current on-demand computing and storage offer attractive opportunities from the perspective of computing resources and infrastructure," she notes. Not only do these services enable organisations to meet their need for increased bandwidth or storage without having to build their own infrastructure, oftentimes these services are faster and cheaper. "However, significant security, privacy and result verifiability concerns are intrinsic to these services, making them outside the reach of organisations who work with sensitive data," Blanton warns.

To understand Blanton's concerns, one only needs to consider how data is handled when it is outsourced to the cloud. Not only are data put outside the control of the owner, but even in the context of best practices, data are likely

SECURELY COMPUTING WITH PRIVATE DATA

OBJECTIVE

To enhance the frontiers of current knowledge for preserving data privacy and integrity through innovative research.

KEY COLLABORATORS

Dr Mehrdad Aliasgari, California State University, Long Beach, USA • **Dr Yihua Zhang**, University of Notre Dame, USA

FUNDING

National Science Foundation (NSF) • Air Force Office of Scientific Research • Air Force Research Laboratory

CONTACT

Dr Marina Blanton
Assistant Professor at the University of Notre Dame

326B Cushing Hall
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, Indiana 46556
USA

T +1 574 631 3637
E mblanton@nd.edu

http://engineering.nd.edu/profiles/mblanton

www.linkedin.com/in/mblanton

http://bit.ly/MarinaBlantonGS

MARINA BLANTON received her PhD from Purdue University in 2007. Her research interests are generally in information security, privacy and applied cryptography. Her recent projects span areas such as secure computation and outsourcing, integrity of outsourced computation and storage, private biometric and genomic computation, privacy-preserving systems for medical and social networks, authentication and anonymity.