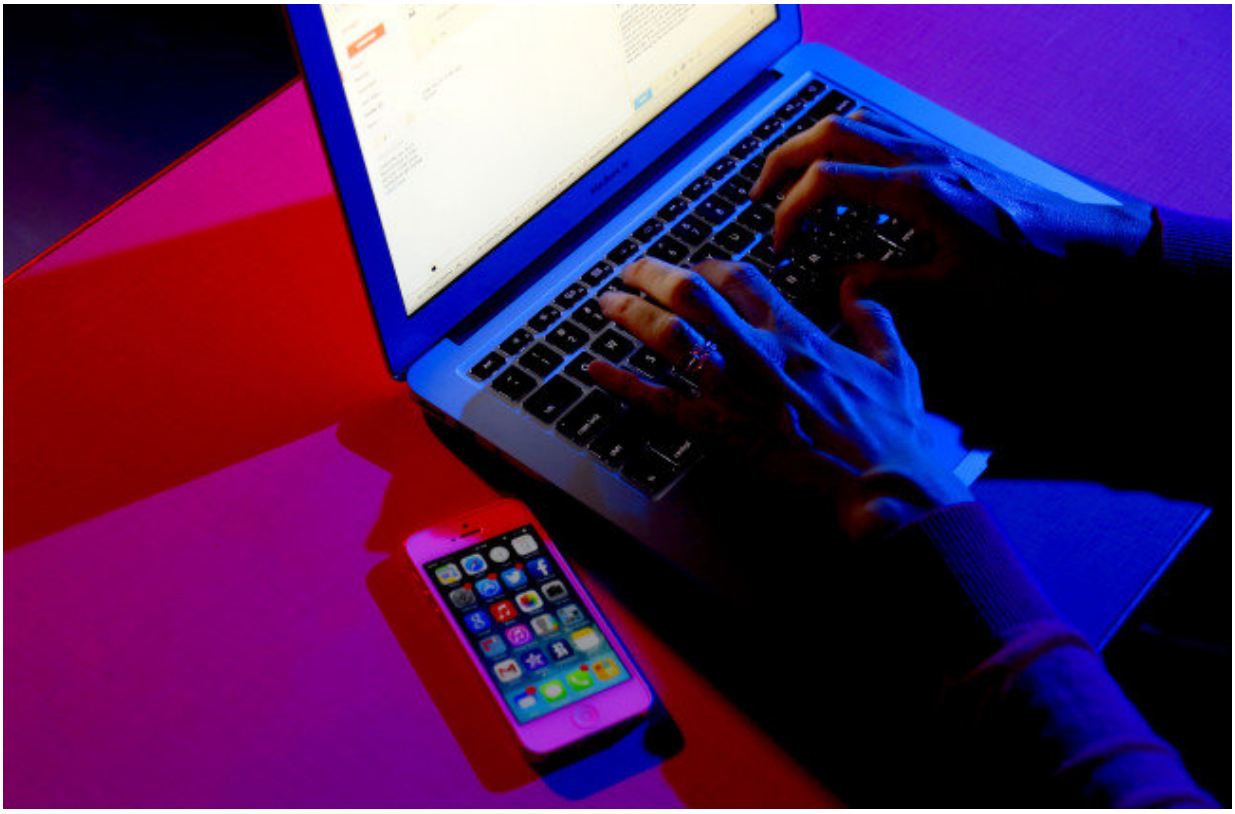
Disguising access patterns to protect sensitive data in the cloud

April 26 2018



Credit: Iowa State University

A team of computer scientists is working to defend against the next potential cyber risk – cloud storage. Wensheng Zhang, an associate professor of computer science at Iowa State University, says cloud users can encrypt sensitive data and information, but how they access the data

may make it vulnerable.

Reports of access pattern-based attacks to cloud storage are rare, Zhang said. Phishing attacks – including a recent hack targeting professors and researchers – are the most common. A 2017 Google study identified as many as 12.4 million potential victims of phishing over the course of a year. However, if hackers can crack the data storage service, Zhang says it is only a matter of time before they try to exploit data access patterns.

"Cloud storage is very convenient, but there are privacy risks," he said. "This kind of threat may be of greater concern to companies or agencies working with very sensitive data. For example, military agencies or some branches of the government."

Here is an example of the threat Zhang is working to prevent: An agency uploads a large dataset to its cloud account. A team analyzing a specific subset of the data regularly accesses the information, creating a pattern. Someone – a rogue employee or hacker who compromised the cloud service – could observe the pattern and make assumptions about the data.

The idea may seem a little farfetched for the average person who uses the cloud to store photos or less sensitive information, but a user storing classified documents or research results in the cloud may feel differently. Zhang says if an agency makes a major decision after accessing that subset of data, hackers can infer the value and focus their efforts on that section, rather than trying to crack the entire file.

Building an efficient solution

Developing the technology to disguise access patterns is technical and complex work. Zhang says the basic premise is to create an algorithm that incorporates a mix of fake and real access requests, making it

difficult to detect a pattern. It sounds simple, but time and cost are two barriers. It needs to be efficient so that the fake access does not delay work or cost too much (bandwidth limitations and cloud service fees), he said.

Zhang; Jinsheng Zhang, lead author and former graduate student in computer science; Qiumao Ma, a graduate student in computer science; and Daji Qiao, an associate professor of electrical and computer engineering, detail one technique in a paper published in the journal Future Internet. The researchers say it is one of the most efficient algorithms proposed for protecting the data access pattern.

The work is ongoing as the team looks for ways to improve performance and efficiency. Zhang says they are also exploring the pros and cons of splitting large datasets across multiple providers, so that access patterns do not reveal the full picture.

"Storage is now more affordable. Five years ago, it was expensive to buy a computer with several hundred gigabytes of storage, but today it is very common," Zhang said. "If users are concerned about privacy, they can keep a small subset of data locally and export the remaining dataset to storage, which can save some cost for protecting the access pattern privacy."

Provided by Iowa State University