



Better Security Through Obfuscation

The quest to find greater security through obscurity.

LAST YEAR, THREE mathematicians published a viable method for hiding the inner workings of software. The paper was a culmination of close to two decades of work by multiple teams around the world to show that concept could work. The quest now is to find a way to make indistinguishability obfuscation (iO) efficient enough to become a practical reality.

When it was first proposed, the value of iO was uncertain. Mathematicians had originally tried to find a way to implement a more intuitive form of obfuscation intended to prevent reverse engineering. If achievable, virtual black box (VBB) obfuscation would prevent a program from leaking any information other than the data it delivers from its outputs. Unfortunately, a seminal paper published in 2001 showed that it is impossible to guarantee VBB obfuscation for every possible type of program.

In the same paper, though, the authors showed that a weaker form they called iO was feasible. While iO does not promise to hide all the details of a logic circuit, as long as they are scrambled using iO, different circuits that perform the same function will leak the same information as each other; an attacker would not be able to tell which



implementation is being used to provide the results they obtain.

“Our motivation in defining the notion of iO was that it escaped the impossibility result for VBB. However, we had no idea if iO could be constructed, and even if it could be constructed, would it be useful for applications,” says Boaz Barak, George McKay professor of computer science in the John A. Paulson

School of Engineering and Applied Sciences at Harvard University, and co-author of the 2001 paper on VBB.

Whatever its utility, for more than a decade iO seemed to be out of reach. A major breakthrough came in 2013, when a team came up with a candidate construction and described a functional-encryption protocol that could be built on top of it. This was quickly fol-

lowed by a slew of proposals for applications that could make use of iO.

One possible application is functional encryption, which makes it possible to selectively hide parts of the same program or data from different users through the use of different decryption keys. This could provide far more fine-grained protection than conventional encryption, where a single key unlocks everything encrypted with it. Other more exotic forms of encryption enabled by iO include deniable encryption, where a user could provide a false key that appears to work but does not reveal information secured by a true key.

Huijia Lin, associate professor in the Paul G. Allen School of Computer Science and Engineering at the University of Washington, points to the possibility of efficient secure multiparty communication, which is difficult to implement using conventional cryptography. “We want multiparty communications, where the overhead to achieve security is so small that it’s as easy as insecure communication. In principle, with iO, you can come up with versions where this is possible.”

The 2013 paper demonstrated a plausible technique for delivering iO, but the novel techniques it employed to obfuscate programs could not guarantee they would not leak too much information. Similar to cryptography, mathematically guaranteed obfuscation relies on mathematical constructs, such as one-way functions, that are practically impossible to reverse without knowledge of the keys used to encode them. An ongoing problem for iO implementors is finding constructs considered secure that, at the same time, provide enough expressive power to transform real-world programs into a form that does not leak information unexpectedly. It was an uphill struggle that took another seven years of work by multiple groups. Often a paper would present a plausible mixture of techniques that would almost as quickly be demonstrated as insufficient to the task.

Amit Sahai, Symantec Chair professor of computer science and director of the Center for Encrypted Functionalities at the University of California at Los Angeles (UCLA), who worked on the 2001 and 2013 papers, says the

cat-and-mouse game of iO constructions being presented and then broken paved the way to a solution. “The process was very important in building our understanding,” he says.

A key breakthrough came last year with the publication of a paper that was the result of a collaboration between Lin, Sahai, and UCLA Ph.D. student Aayush Jain, which was based on assumptions they consider to be well-founded, though some of them are novel in the field of cryptography. “We showed how to construct iO from problems that have been around for at least a decade,” Sahai says.

The paper rests on the assumed security of four mathematical problems that the authors claim have well-established histories. Some, such as problems based on the elliptic curves used in cryptography, have been widely used. They also found a technique that has not been heavily explored cryptographically, but which seems to offer a high degree of protection against information leakage. Mathematician Richard Hamming proposed the idea of random linear codes for error cor-

ACM News

Semantics Beats Syntax

IBM, Google, and Microsoft all are poised to release semantic engines (algorithms using the meaning of words) to supplement their current syntax engines (using the spelling of words). Their common goal is to extend their natural language processing (NLP) capabilities into engines that rival human semantics (our understanding of what language, words/sentences, mean).

Today’s syntax-only engines are blind to the meaning of keywords used to ascertain results. A human understands that “where Alan Turing was born” means the same as “the birthplace of Alan Turing” and “the town where Alan Turing was delivered as a baby.” Their syntax differs, but each phrase’s meaning, or semantics, are identical (“London” is the answer to all three). People understand this immediately, but computers—not so much.

Consequently, all three companies are developing algorithms that understand the

meaning of words. Google and Microsoft are both building semantic engines that add metadata to sentences (Google) or words (Microsoft) using clusters of processors running multiple deep neural networks (called transformers, which use massive parallelization).

IBM

For its semantic engine, IBM chose to augment neural networks with symbolic logic, reducing the number of examples it requires to learn. Said Forrester Research principal analyst Kjell Carlsson, “IBM’s semantics uses a much more efficient encoding of knowledge, enabling high performing enterprise use-cases to be built with significantly smaller training examples.”

In addition, said Carlsson, “IBM’s neuro-symbolic approach enables higher accuracy with less training data, plus it also enables engineers to ‘teach’ a model logical relationships that domain experts know to be true, which

is far more efficient than having these relationships be learned by transformers.”

Google

Google and Microsoft have both released free test versions of their semantic transformers. Google’s, called Semantic Experiences, tackles four separate application domains, plus a roll-your-own capability.

Google’s demos include “Verse-by-Verses,” a semantic “experience” that composes poetry; “Talk-to-Books,” which answers queries based on statements found in current books; “Semantris,” a word-association game, and the free-form “Create Your Own Semantic Experience” tool.

Microsoft

Microsoft aims to release the first semantic-based commercial product. Using word-level granularity in meaning encoding works best, according to Microsoft’s Luis Cabrera-Cordon,

a group program manager for Azure, who describes Microsoft’s “semantic search on Azure [as offering] the best combination of search relevance, developer experience, and cloud service capabilities.”

Forrester Research’s Carlsson said, “The biggest recent advancements in AI have been in (deep) learning, which has opened up the world of unstructured data (vision, text, voice, logs) for analysis at scale, but what we really want is both learning and knowledge. Learning enables us to update and acquire new knowledge, and knowledge makes learning more efficient, governable, and valuable. What makes these new deep learning-infused semantic methods exciting is their potential to deliver both, dramatically expanding not just NLP, but all machine learning use-cases.”

—R. Colin Johnson is a Kyoto Prize Fellow who has worked as a technology journalist for two decades.

rection 70 years ago, in which a message is encoded using a matrix of values that are generated randomly. Since then, scientists have searched unsuccessfully for an efficient way to reverse the process without knowledge of how the data was encoded. That, in turn, led to the conjecture that performing such decoding efficiently is hard, and that the constructs could be employed to help build iO implementations. Sahai stresses that to defeat this conjecture, it would take a mathematical breakthrough that has eluded communication scientists for decades.

Though it rests on assumptions that are on firm footing, a stumbling block of the Jain, Lin, and Sahai proposal is the complexity of the construction. “This is just the first construction where the pieces finally connected to form a secure scheme. But we have all these steps, all these transformations we have to make to the program to obfuscate it,” says Sahai. “Each step introduces a huge overhead.”

Estimates of some older work on iO illustrate the computational complexity gap that researchers need to bridge. One paper published in 2016 based on techniques now considered insecure showed that even a simple 80-bit point function that is zero for all inputs except one would consume more than 10GB of memory and take three minutes to execute. Sahai says the overhead of their current scheme is so high, it is not practical to even estimate it.

Barak says the computational overhead is unlikely to be insurmountable. “In cryptography, we’ve had examples, such as multiparty secure computation and probabilistically checkable proofs, where the initial constructions were almost comically inefficient, but over time people have improved them by 20 or so orders of magnitude,” he says.

Though computational overhead is an issue and may mean practical applications will not appear for over a decade, Lin says it is equally important to create a construction using fewer or simpler assumptions. A more compact approach would improve confidence in iO as a building block for secure computation. Numerous groups are now looking to see what can be distilled out of the existing work to create a construction that is

“Often you go back and realize you didn’t need certain steps. We are in that tightening mode, and also alternative-finding mode, but we just don’t know how long it will take.”

conceptually simpler.

“Often you go back and realize you didn’t need certain steps. We are in that tightening mode, and also alternative-finding mode,” Sahai says. “But we just don’t know how long it will take.”

The main question to be answered in future constructions is which assumptions will provide a way forward for reducing complexity and overhead while ensuring acceptance of iO is on a firm footing.

One approach that has been taken by multiple groups over the past couple of years is to try to let an iO construction rest on a single-core mathematical problem. A major candidate for that is the Learning With Errors (LWE) problem developed by computer scientist and mathematician Oded Regev more than a decade ago, for which he was awarded the 2018 Gödel Prize.

LWE already forms the basis for lattice-based cryptographic systems and is being actively pursued because it is generally considered to be safe from attack by quantum computing. Barak says LWE has conceptual similarities to the random linear codes used in the work published by Jain, Lin, and Sahai, which makes it seem a viable approach.

Yet that is not necessarily the path iO will take.


Jain says the direction taken in work following the 2020 paper has led him and his colleagues to focus instead on building iO without LWE. Lin points out that noise that helps LWE maintain security in conventional cryptography leaks information that could compromise iO. Lin says the assumption that

underpins the random linear codes seems to have peculiar properties that are worth exploring further.

Though there may be skepticism in the cryptographic community about the more-novel assumptions that iO seems to rely on, at least for the time being, some of that may simply be attributed to their relative novelty. Lin points to the way that papers from several decades ago used to justify why they used Diffie-Hellman key exchange, whereas today it is widely accepted.

“People’s confidence in an assumption tends to grow over time, and with the number of papers that use the assumption,” Lin says.

Sahai believes further work may revive the concept of VBB obfuscation. “The impossibility result was overinterpreted by the community at large,” he says, on the basis that though the 2001 paper ruled out VBB obfuscation for all software, many common forms of software could yet be candidates. That could in turn make it possible to run software on untrusted machines without fear of the code being reverse-engineered from its code.

Such applications, including those of iO, likely lie some distance into the future. 

Further Reading

Barak, B., Goldreich, O., Impagliazzo, R., Rudich, S., Sahai, A., Vadhan, S., and Yang, K. On the (Im)possibility of Obfuscating Programs *Advances in Cryptology* (2001). 2010 revision: <https://www.boazbarak.org/Papers/obfuscate.pdf>

Garg, S., Gentry, C., Halevi, S., Raykova, M., Sahai, A., and Waters, B.

Candidate Indistinguishability Obfuscation and Functional Encryption for all circuits 54th Annual Symposium on Foundations of Computer Science (FOCS), (2013)

Jain, A., Lin, H., and Sahai, A. Indistinguishability Obfuscation from Well-Founded Assumptions *IACR Cryptology ePrint Archive*: 1003 (2020) <https://eprint.iacr.org/2020/1003>

Brakerski, Z., Döttling, N., Garg, S., and Malavolta, G.

Candidate iO From Homomorphic Encryption Schemes *IACR Cryptology ePrint Archive*: 394 (2020) <https://eprint.iacr.org/2020/394>

Chris Edwards is a Surrey, U.K.-based writer who reports on electronics, IT, and synthetic biology.

© 2021 ACM 0001-0782/21/8 \$15.00