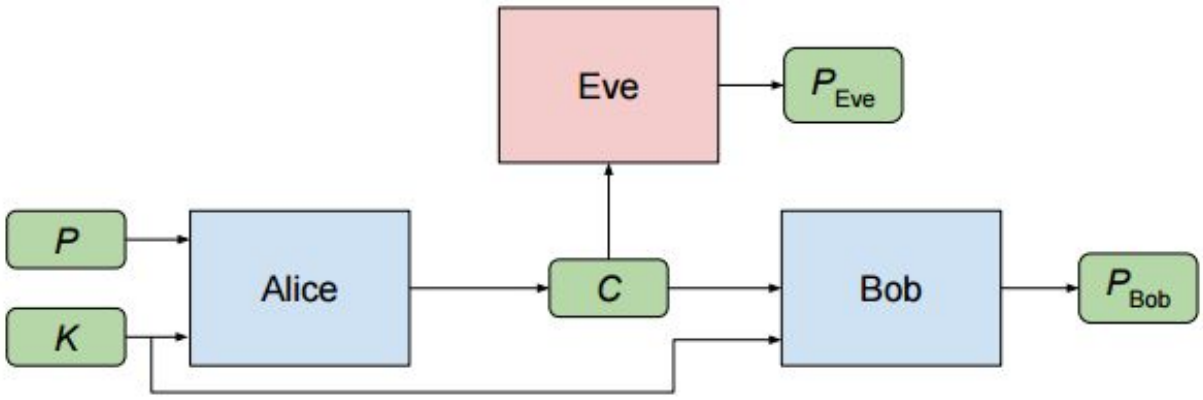
Neural networks learn more about protecting communications: Just ask Alice, Bob and Eve

October 30 2016, by Nancy Owano



Alice, Bob, and Eve, with a symmetric cryptosystem. Credit: arXiv:1610.06918 [cs.CR]

(Tech Xplore)—What will Google think of next? Actually, if you are looking at the goings on of the Google Brain team you are right to suspect they don't want to be thinking at all. They would rather put their AI resources to work.

They are interested in what machines can come up with for encryption methods. In other words, the Google team's neural networks are playing with their own encryption, working out use of their own techniques.

On Friday, Mike Wehner in *The Daily Dot* said, "Google engineers

decided to let a trio of AI minds put their own encryption skills to the test."

Alice, Bob and Eve were the minds and each was given a specific goal.

Alice, send Bob a message. (Both Alice and Bob were given matching keys with which to encode and decode their conversation.)

Eve, intercept and decode it. (Eve had to try to translate the encrypted message into plain text without the key.)

None of the three had it easy. Wehner wrote that "Alice had to come up with her own encryption algorithm, but neither Bob nor Eve was given that information. Bob didn't have any idea how the key should be applied to the encrypted message, and Eve was basically working completely from scratch."

Timothy Revell in *New Scientist* also described what was going on.

"To make sure the message remained secret, Alice had to convert her original plain-text message into complete gobbledygook, so that anyone who intercepted it (like Eve) wouldn't be able to understand it. The gobbledygook – or "cipher text" – had to be decipherable by Bob, but nobody else. Both Alice and Bob started with a pre-agreed set of numbers called a key, which Eve didn't have access to, to help encrypt and decrypt the message."

Wehner summarized what the test results indicate: AI can decode encrypted text with little or no guidance. On the other hand, it is not hard to fool a blind AI forced to guess how to solve the problem. Wehner said, "In short, AI is great at encryption, but not so great at blind decryption, at least for now."

Nathaniel Mott in *Inverse* said such research shows that AI is getting better at teaching itself. "This complements other efforts to make AI smarter," he added.

Revell in *New Scientist* meanwhile reminded us that "Computers have a very long way to go if they're to get anywhere near the sophistication of human-made encryption methods."

What's next? Mott wrote that "Google's researchers said the next step for Alice, Bob, and Eve could involve other cryptographic protections like pseudorandom number generations or steganography. They don't expect A.I. to ever become a codebreaker, but they could at least help us keep our communications private, and analyze the metadata associated with digital messages."

The authors stated in conclusion that "While it seems improbable that neural networks would become great at cryptanalysis, they may be quite effective in making sense of metadata and in traffic analysis."

The Google Brain two-author study submitted to the arXiv is titled "Learning to protect communications with adversarial neural cryptography" by Martın Abadi and David Andersen. They wrote, "We ask whether neural networks can learn to use secret keys to protect information from other neural networks."

More information: Learning to Protect Communications with Adversarial Neural Cryptography, arXiv:1610.06918 [cs.CR] arxiv.org/abs/1610.06918

Abstract
We ask whether neural networks can learn to use secret keys to protect information from other neural networks. Specifically, we focus on ensuring confidentiality properties in a multiagent system, and we

specify those properties in terms of an adversary. Thus, a system may consist of neural networks named Alice and Bob, and we aim to limit what a third neural network named Eve learns from eavesdropping on the communication between Alice and Bob. We do not prescribe specific cryptographic algorithms to these neural networks; instead, we train end-to-end, adversarially. We demonstrate that the neural networks can learn how to perform forms of encryption and decryption, and also how to apply these operations selectively in order to meet confidentiality goals.