

New theory could yield more-reliable communication protocols

December 12 2014, by Larry Hardesty



Credit: Jose-Luis Olivares/MIT

Communication protocols for digital devices are very efficient but also very brittle: They require information to be specified in a precise order with a precise number of bits. If sender and receiver—say, a computer and a printer—are off by even a single bit relative to each other, communication between them breaks down entirely.



Humans are much more flexible. Two strangers may come to a conversation with wildly differing vocabularies and frames of reference, but they will quickly assess the extent of their mutual understanding and tailor their speech accordingly.

Madhu Sudan, an adjunct professor of electrical engineering and <u>computer science</u> at MIT and a principal researcher at Microsoft Research New England, wants to bring that type of flexibility to computer <u>communication</u>. In a series of recent papers, he and his <u>colleagues</u> have begun to describe theoretical limits on the degree of imprecision that communicating computers can tolerate, with very real implications for the design of communication protocols.

"Our goal is not to understand how human communication works," Sudan says. "Most of the work is really in trying to abstract, 'What is the kind of problem that human communication tends to solve nicely, [and] designed communication doesn't?'—and let's now see if we can come up with designed communication schemes that do the same thing."

One thing that humans do well is gauging the minimum amount of information they need to convey in order to get a point across. Depending on the circumstances, for instance, one co-worker might ask another, "Who was that guy?"; "Who was that guy in your office?"; "Who was that guy in your office this morning?"; or "Who was that guy in your office this morning with the red tie and glasses?"

Similarly, the first topic Sudan and his colleagues began investigating is compression, or the minimum number of bits that one device would need to send another in order to convey all the information in a data file.

Uneven odds

In a paper presented in 2011, at the ACM Symposium on Innovations in



Computer Science (now known as Innovations in Theoretical Computer Science, or ITCS), Sudan and colleagues at Harvard University, Microsoft, and the University of Pennsylvania considered a hypothetical case in which the devices shared an almost infinite codebook that assigned a random string of symbols—a kind of serial number—to every possible message that either might send.

Of course, such a codebook is entirely implausible, but it allowed the researchers to get a statistical handle on the problem of compression. Indeed, it's an extension of one of the concepts that longtime MIT professor Claude Shannon used to determine the maximum capacity of a communication channel in the seminal 1948 paper that created the field of information theory.

In Sudan and his colleagues' codebook, a vast number of messages might have associated strings that begin with the same symbol. But fewer messages will have strings that share their first two symbols, fewer still strings that share their first three symbols, and so on. In any given instance of communication, the question is how many symbols of the string one device needs to send the other in order to pick out a single associated message.

The answer to that question depends on the probability that any given interpretation of a string of symbols makes sense in context. By way of analogy, if your co-worker has had only one visitor all day, asking her, "Who was that guy in your office?" probably suffices. If she's had a string of visitors, you may need to specify time of day and tie color.

Existing compression schemes do, in fact, exploit statistical regularities in data. But Sudan and his colleagues considered the case in which sender and receiver assign different probabilities to different interpretations. They were able to show that, so long as protocol designers can make reasonable assumptions about the ranges within



which the probabilities might fall, good compression is still possible.

For instance, Sudan says, consider a telescope in deep-space orbit. The telescope's designers might assume that 90 percent of what it sees will be blackness, and they can use that assumption to compress the image data it sends back to Earth. With existing protocols, anyone attempting to interpret the telescope's transmissions would need to know the precise figure—90 percent—that the compression scheme uses. But Sudan and his colleagues showed that the protocol could be designed to accommodate a range of assumptions—from, say, 85 percent to 95 percent—that might be just as reasonable as 90 percent.

Buggy codebook

In a paper being presented at the next ITCS, in January, Sudan and colleagues at Columbia University, Carnegie Mellon University, and Microsoft add even more uncertainty to their compression model. In the new paper, not only do sender and receiver have somewhat different probability estimates, but they also have slightly different codebooks. Again, the researchers were able to devise a protocol that would still provide good compression.

They also generalized their model to new contexts. For instance, Sudan says, in the era of cloud computing, data is constantly being duplicated on servers scattered across the Internet, and data-management systems need to ensure that the copies are kept up to date. One way to do that efficiently is by performing "checksums," or adding up a bunch of bits at corresponding locations in the original and the copy and making sure the results match.

That method, however, works only if the servers know in advance which bits to add up—and if they store the files in such a way that data locations correspond perfectly. Sudan and his colleagues' protocol could



provide a way for servers using different file-management schemes to generate consistency checks on the fly.

"I shouldn't tell you if the number of 1's that I see in this subset is odd or even," Sudan says. "I should send you some coarse information saying 90 percent of the bits in this set are 1's. And you say, 'Well, I see 89 percent,' but that's close to 90 percent—that's actually a good protocol. We prove this."

"This sequence of works puts forward a general theory of goal-oriented communication, where the focus is not on the raw data being communicated but rather on its meaning," says Oded Goldreich, a professor of computer science at the Weizmann Institute of Science in Israel. "I consider this sequence a work of fundamental nature."

"Following a dominant approach in 20th-century philosophy, the work associates the meaning of communication with the goal achieved by it and provides a mathematical framework for discussing all these natural notions," he adds. "This framework is based on a general definition of the notion of a goal and leads to a problem that is complementary to the problem of reliable communication considered by Shannon, which established <u>information theory</u>."

More information: "Communication with Imperfectly Shared Randomness": <u>people.csail.mit.edu/madhu/pap ... 14/comm-isr-full.pdf</u>

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology



Citation: New theory could yield more-reliable communication protocols (2014, December 12) retrieved 7 May 2024 from <u>https://phys.org/news/2014-12-theory-yield-more-reliable-protocols.html</u>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.