Zero-sum game

The future of humanity depends on our ability to navigate the perils of the nuclear age, and especially to steer around the potential catastrophe of actual nuclear warfare. Things may soon get trickier, as climate change may bring heightened conflict over energy and other resources, and as dramatic political changes seem to be ushering in a new era of shifting international power balances. Much could depend on fragile nuclear arms limitation agreements, not only between the United States and Russia in the near term, but between many other nuclear nations in the future as well.

Our safety over the past seven decades has rested on the balance of nuclear forces, with no party ever gaining such a technological edge that it might believe it could win through a unilateral strike. Until now, verification of this balance has been made through counts of associated delivery hardware — planes, submarines and missiles. Yet the next era's agreements may require verification of the true identity of actual nuclear devices, and may run up against a subtle challenge: how to enable inspectors to verify the identity of two warheads, without learning anything about the sensitive classified details of the warheads' materials or structures. Imagine, for example, that a nation deployed a new device, claiming it as just another copy of an existing design. Other nations might doubt that, and want to make sure.

Could secure verification of this kind be possible? Well, yes, at least in principle. One idea currently in play is to use automated systems to carry out the analysis of the two devices — say, through neutron transmission and scattering experiments, which can offer unique identifying images. The analysers could be programmed to report only the ultimate yes–no answer, while keeping all other information hidden, even destroying it in the process. Systems of this kind have been developed by national laboratories in the US, UK and in Russia.

Yet computational systems and software are notoriously prone to tampering, and experts wonder if this idea could ever be sufficiently trusted in the nuclear setting. As a result, scientists have recently been exploring other ideas, one of which is to use the notion of so-called zero-knowledge proofs. These are techniques, developed in computer science, for proving that a statement is true without anyone involved in the process gaining information on

Computational systems are prone to tampering and experts wonder if they can ever be trusted in the nuclear setting.

why it might be true. This kind of thing is required, for example, in building secure voting systems, and the ideas have recently been adapted to the physical world in closely related zero-knowledge protocols.

Two years ago, for example, Alex Glaser, Boaz Barak and Robert Goldston proposed one conceptual scenario (*Nature* **510,** 497–502; 2014). Imagine that an inspector, sent to test the identity of the new and old devices, has access to the new device and a trusted example of the reference warhead. Neutron transmission and scattering data for both new and reference warheads would reveal differences, or prove that they are identical, but also reveal classified information to the inspector. To avoid this, Glaser *et al.* suggested that the inspection proceed as follows. The nation should submit to the inspector a set of, say, two 'preloads' for the neutron detectors — these are partial exposures of the detectors carrying the negative image of the radiograph of the true item. In the protocol, the inspector selects a preload at random, and irradiates one of the objects, again selected randomly, and notes the output.

If the nation is playing straight, and sets both preloads to be the precise negative of the actual exposure patterns that would result from irradiation of the true older warhead design, then the detectors will end up all having the same exposure, showing a perfect match, while revealing no information. Of course, the host might try to cheat by submitting a preload that is altered and designed to match a similarly modified object. But the random element of the protocol prevents this. Because the inspector chooses both the preload to use and which object to test at random, there's only a 50% chance of getting away with such trickery in any trial. Run a small number of trials, and it becomes virtually certain that any subterfuge will be detected, even while ensuring that no actual information about the objects is revealed.

In their initial proposal, Glaser *et al.* only described the scheme and illustrated it with simulations, but the group has now gone further and demonstrated the idea in an actual physical experiment (S. Philippe *et al.*, *Nat. Commun.* **7,** 12890; 2016). This appears to be the very first time a zero-knowledge proof of physical properties has been demonstrated. The group tells me that they're now running additional experiments with uranium and plutonium objects of various types. Meanwhile, other researchers are exploring interesting variations.

One of these isn't a zero-knowledge protocol per se, but achieves much the same result, with the secrecy of revealed information protected by a physically-encoded cryptographic key (R. Scott Kemp *et al.*, *Proc. Natl Acad. Sci. USA* **113,** 8618–8623; 2016). Here the idea is that an inspector would form a three-dimensional (3D) tomographic image of, say, the new warhead, by irradiating the object with high energy photons and detecting the nuclear resonance fluorescence generated by the 3D distribution of isotopes within. By constructing the tomographic images using exposures at randomly selected orientations, the inspector should reveal any trickery by the host nation. The inspector then also images the reference item, to see if the images of the two objects match.

However, to avoid the inspector seeing information in any of the images, these get passed through a thin scattering foil before being available, which scrambles the image in what is effectively a cryptographic transform determined by the detailed microscopic and highly random structure of the foil. Unscrambling would require detailed knowledge of the foil structure — and the host nation supplies the foil and keeps this structure secret.

Hence, again, whether the objects are identical or not can be verified, without the inspector learning anything about the objects in question. No doubt these proposals only reflect the beginning of what might be possible along these lines. And, of course, nothing will probably ever be totally fool-proof — each technique, for example, assumes that the reference item can be trusted as authentic. But solving part of the problem would be a significant step forward in engineering a means to continuing that delicate nuclear balance on which our future safety depends. ❐

MARK BUCHANAN

Corrected online: 15 December 2016

Correction

In the Thesis 'Zero-sum game' (*Nature Physics* **12,** 1084; 2016), the citation to Philippe *et al.* was incorrect and should have read '*Nat. Commun.* **7,** 12890; 2016'. This has been corrected after print 15 December 2016.