SCIENCE & GLOBAL SECURITY
2020, VOL. 28, NO. 1, 48–59
https://doi.org/10.1080/08929882.2020.1728885

Routledge
Taylor & Francis Group

Physical Public Templates for Nuclear Warhead Verification

Alexander Glaser[a], Boaz Barak[b], Moritz Kütt[c] and Sébastien Philippe[a]

[a]Program on Science and Global Security, Princeton University, Princeton, NJ, USA; [b]Harvard School of Engineering and Applied Sciences, Cambridge, MA, USA; [c]Institute for Peace and Security Research, Hamburg, Germany

ABSTRACT

Passive gamma spectroscopy has been successfully used for nuclear warhead inspection systems based on the template-matching approach. The most prominent example of such a system is Sandia's Trusted Radiation Identification System (TRIS), which is based on an earlier system used at Pantex since 1994 to confirm the identities of containerized plutonium pits. Remarkably, TRIS uses only 16 energy bins, i.e., 16 numbers, to accomplish this task. Additional experiments have shown that such a template-matching method could be performed in a way that does not reveal classified information. To be used in a real inspection setting, however, inspectors must gain confidence that the system hardware and software work as designed and display genuine measurements through a process known as authentication. It also requires establishing and maintaining confidence in the template, i.e., that the data characterizing the treaty accountable item is genuine and has not been altered. In the case of TRIS, the template data are stored electronically and signed as a whole, such that no information about the template can ever be shared with inspectors as a confidence-building measure. Here, we propose an inspection protocol that uses a different approach: Information is stored in the form of punched cards that encode the secret template. Public masks can be used to reveal selected features of the template, e.g., total counts in particular energy bins, while keeping others secret, constraining certain physical properties of the treaty accountable item and providing increasing levels of transparency. We illustrate our approach using Princeton's Information Barrier Experimental II based on a vintage 6502 processor.

Background

Template-matching approaches based on passive gamma spectroscopy combine several highly desirable features for inspection systems used for warhead confirmation measurements. They are simple to set up and use in the

CONTACT Alexander Glaser alx@princeton.edu Program on Science and Global Security, Princeton University, 221 Nassau St, 2nd floor, Princeton, NJ, 08540, USA

field without raising safety concerns. Moreover, unlike inspection systems based on the attribute approach, they can work with low-resolution detectors that are robust and can provide some inherent information security. The most prominent example of such a system is Sandia's Trusted Radiation Identification System (TRIS), which is based on an earlier system used at Pantex since 1994 to confirm the identities of containerized pits. Sandia researchers believe that "template matching can be performed in a way that is robust and does not reveal classified information."[1] While there are strategies to defeat passive detection systems,[2] these systems can establish a first confirmation layer that can be followed by additional, more involved measurements, perhaps for randomly selected items, later on.

Despite the strengths of passive template-matching, no information barrier system has so far been successfully authenticated.[3] This is mostly due to the difficulty of establishing the inspector's trust in the hardware and software, including the integrity of the template data. To address these concerns, at least partially, we offer two main ideas. First, we examine the possibility of storing secret data, i.e., the template, on non-electronic physical objects to make the hardware authentication process simpler. In particular, we use traditional punched cards to store the data.[4] In our case, data considered sensitive are never stored permanently on electronic media; they only exist in volatile memory when the inspection system is powered and initialized. Second, we also propose a non-electronic scheme to selectively make public information about the template as part of possible transparency and confidence-building measures, for example, by revealing specific attributes of the template or information relative to the calibration of the apparatus. To do so, we introduce masks that can be placed on top of our template punched card to release bitwise information. Overall, this effort is part of our "vintage verification" project, which explores the potential of hardware from the 1970s as platforms for trusted computing.[5]

Inspection concept and protocol

Similar to Sandia's TRIS, which uses only 16 numbers representing counts per energy bins to generate the template gamma spectrum, we also limit the size of the template. For this proof-of-concept, we use 12 bins, each 250 keV wide, covering the entire energy range from 0 keV to 3,000 keV. In general, just as in the case of TRIS, the bins could be of different widths, there could be gaps between them, and they could be weighted or combined in different ways. In practice, the host and inspectors would need to agree on the details prior to inspection. The format of the punched card proposed below allocates 24 bits per bin, i.e., it can store more than 16 million counts per channel. For simplicity, we assume that the system acquires

exactly $2^{18} = 262{,}144$ counts before terminating the measurement.[6] For a count rate of about 2,000 counts per second, typical for the measurements on plutonium objects describe here, a spectrum is acquired in 2 to 3 minutes, making the confirmation process relatively fast and practical. The inspection protocol has a private phase and a public phase, which are summarized in the following.

Private phase

Generating the secret template

The host has at least one warhead of the relevant type available and can use this item to determine its gamma spectrum with excellent accuracy using the same type of measurement equipment that will also be used during future inspections.

Based on this spectrum, the host can then determine the numerical values for the 12 broad-energy bins that will serve as the template. As an example, Figure 1 shows the reference spectrum and the respective numerical values of the template of a plutonium ball measured at the Device Assembly Facility in August 2017. The system used for these measurements is based on a sodium-iodide detector and uses a number of thoriated

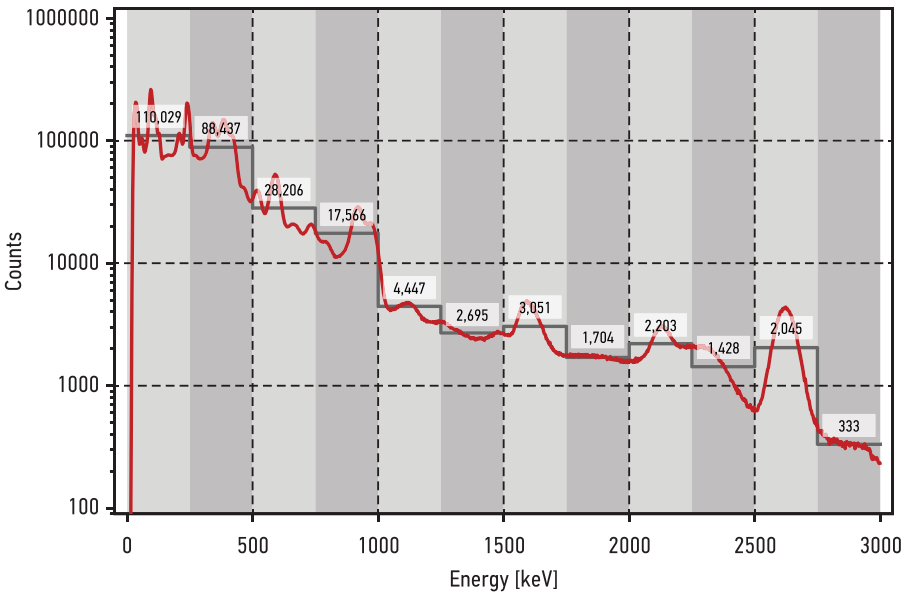

Figure 1. Sample gamma radiation spectrum from a 4.5-kg solid plutonium ball measured at the Device Assembly Facility in August 2017. The spectrum includes signatures from a weak calibration source (here thorium with its characteristic peak at 2.614 MeV.) Also shown is the 12-bin representation of this spectrum, including the 12 numerical values that make up the template and would typically be considered highly sensitive in the case of a nuclear warhead. The data are scaled down to a total of $2^{18} = 262{,}144$ counts.

welding rods that surround the detector crystal to "self-calibrate" the detector in regular intervals.[7]

Making the punched card encoding the secret template

Once the host has privately determined the numerical values that describe the signature from a particular warhead or warhead-component type, the template card can be prepared. Figure 2 shows the design that we chose for our prototype system. We envision that a total of 512 bits are available on the card in order to make brute-force attacks based on the card's SHA3-512 hash (discussed further below) impossible. We allocate 24 bits per energy bin, adding up to 288 bits of template data. Another 208 bits are available for padding bytes, which can be arbitrarily chosen by the host. The final 16 bits on the card are used for a standard (Fletcher 16) checksum to ensure that the card has been properly read.[8]

Public phase

Validating the template card

Before any meaningful inspections can be carried out, the template card must be validated. To do this, a trusted reference item, i.e., a nuclear warhead of a specified type, has to be available. The inspector could select this reference item, perhaps directly from the front section of a ballistic missile. Figure 3 illustrates the key steps of this phase.

To begin the validation process, the template card is read with a punched card reader and its cryptographic hash is calculated and displayed by the inspection system. At this point, the use of a punched card offers an important advantage over electronic storage of the data. A template reader would normally be a separate electronic device that would need to be

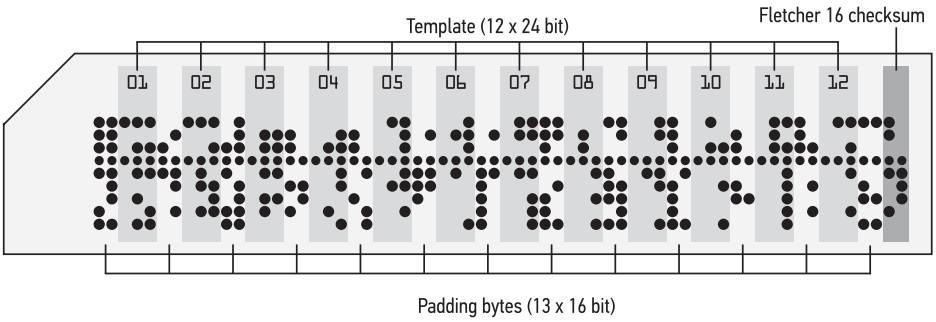

Figure 2. Secret punched card for the 12 template values shown in Figure 1. Twenty-four bits are available for each bin of the template, for a total of 288 bits of spectrum data. Arbitrarily chosen padding bytes provide extra entropy. The last 16 bits are a checksum to ensure that the data have been read in correctly. Overall, the card has a storage capacity of 512 bits (or 64 bytes).

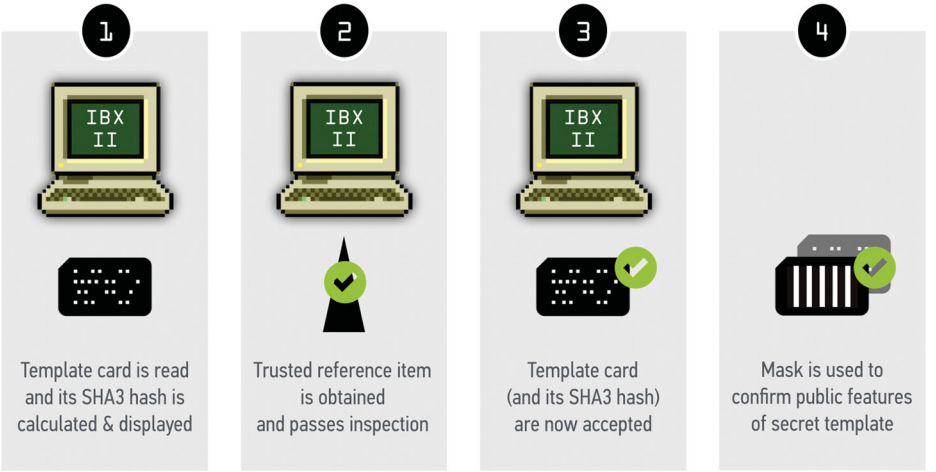

Figure 3. Steps of the protocol to validate the template card with a trusted reference item and to possibly and selectively confirm public information about the template using a dedicated mask.

certified and authenticated. A punch card can be read with a simple device, such as an electromechanical reader, which would be relatively easy to authenticate. After the template is read, its cryptographic hash is calculated and displayed by the inspection system, such that the host commits to the template data and cannot modify it.[9] As an example, the SHA3-512 hash of the card shown in Figure 2 is:

```
3d05e4f4e1de1619a4075c66d91076da
93ebd94f4c9251db1c9e1854afae0498
1666836f92502a30b83eaeafa383d5d5
58c6d00ad149a0f3d51002b1d0e7a8fd
```

Using SHA3-512, the hash has 512 bits,[10] just like the message encoded on the card. The group of SHA3 functions is considered to be among the strongest hash functions available today. For a cryptographic hash function to be viable, it has to be very hard to find a valid message based on a known hash, to find a second message that produces the same hash as a given message, and to find two arbitrary messages that produce the same hash. In our context, the hash function ensures that the inspector cannot extract the template from the public hash and that the host cannot modify the template without also changing its hash.

The public hash of the template card will be used in all future inspections to confirm that the host is using the same card that was used during this validation phase[11] and that the template has not been modified.

The system is now initialized and ready for measurements. If the trusted reference item passes the inspection, then the template card and its public hash are considered "valid" and accepted for future use. In the case of the IBX II,[12] we use a standard χ^2-test to compare the measured values (N_i) with the values of the secret template (T_i):[13]

$$\chi^2 \approx \sum_{i=1}^{12} \frac{(N_i - T_i)^2}{T_i}$$

For the 11 degrees of freedom of this problem,[14] the χ^2-value obtained for a valid item is greater than 30 in one out of 1,000 cases, i.e., a valid item is incorrectly flagged as invalid with a probability of about 0.1%. It is also worth noting that expected variations between warheads of the same type (due to manufacturing tolerances, variations in material composition and age, etc.) would affect the outcome of the χ^2-test and broaden the distribution of χ^2-values. In general, the threshold value for the χ^2-test would be subject to negotiation between the parties, and agreed strategies for repeated testing could be used to reduce the overall false positive rate.[15]

Inspecting candidate items

To prepare for the inspection of a candidate item, the system is initialized with the secret template card provided by the host. The system displays the SHA3-512 hash of the template, and the inspector can confirm that it is identical to the public value previously recorded. Once the measurement is complete and the inspection system displays a "pass" signal (versus a "fail" signal) based on the χ^2-test, then the item is accepted as a valid item. Figure 4 shows the result of 100 repeated measurements of the plutonium ball discussed above. In this example, all χ^2-values are significantly below the notional threshold of $\chi^2_T = 30$.

Revealing selected template data (optional)

The punched card method described here has the important advantage of allowing to reveal some information about the template and therefore increasing the confidence in the overall procedure. As part of the protocol, the host and inspector can agree on using a mask that will be placed on top of the template card during the template validation process (Step 4 in Figure 3). The mask will reveal selected features of the template data to the inspector, while concealing the rest of the card. Figure 5 shows a few examples. There are two important benefits of this approach. The first is the possibility to reveal information that would complement the authentication of the inspection hardware by providing specific information on the measurement, for example, the total count and the presence of a calibration

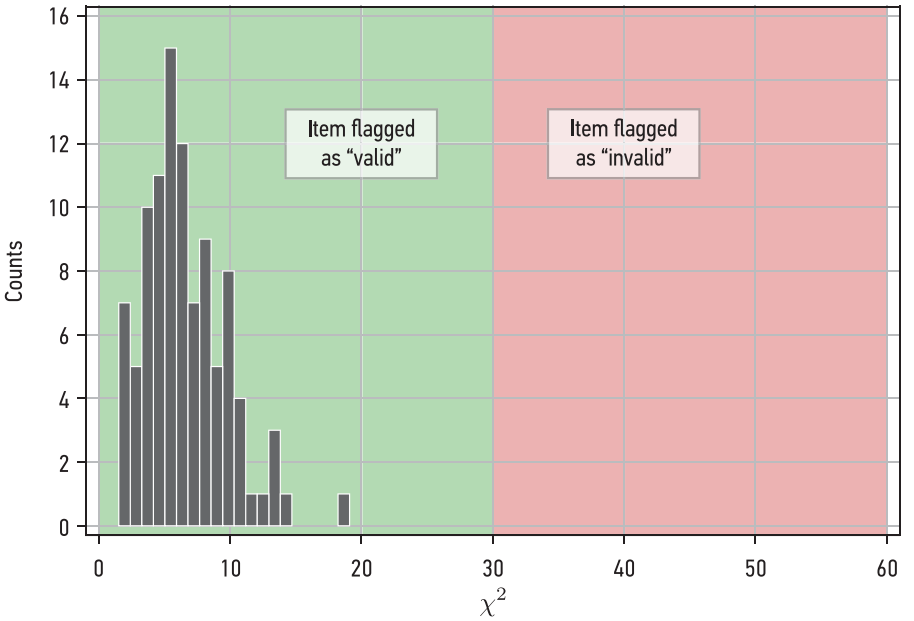

Figure 4. Distribution of χ^2-values for 100 repeated measurements of the 4.5-kg solid plutonium ball. Acquired data are compared against the template values listed in Figure 1. An item passes the test if the χ^2 is below a previously agreed threshold. Here, for a notional threshold value of $\chi^2_T = 30$, all items pass with a good margin.

peak in a specific bin. The second is the possibility to reveal *a posteriori* specific features of the template to help with the disarmament process. This is an important advantage of this methods. Something like that is not possible if the entire template data are stored electronically and signed as a whole.

Toward public templates

In general, the gamma radiation signature of a nuclear warhead is considered highly sensitive information. The purpose of information barriers is to protect this information, but authentication of these devices has proven elusive. As briefly discussed below, confirming correct operation of the equipment may be less challenging if some features of the template were made public. In fact, if the entire template was considered non-sensitive, no information barrier would be needed.[16]

Without assessing the desirability or likelihood of weapon states publicly revealing selected features of warhead radiation signatures, here, we explore how such a process could be implemented in a selective or gradual manner.

The proposed approach envisions separating secret and public information about the template using dedicated masks, which enables some

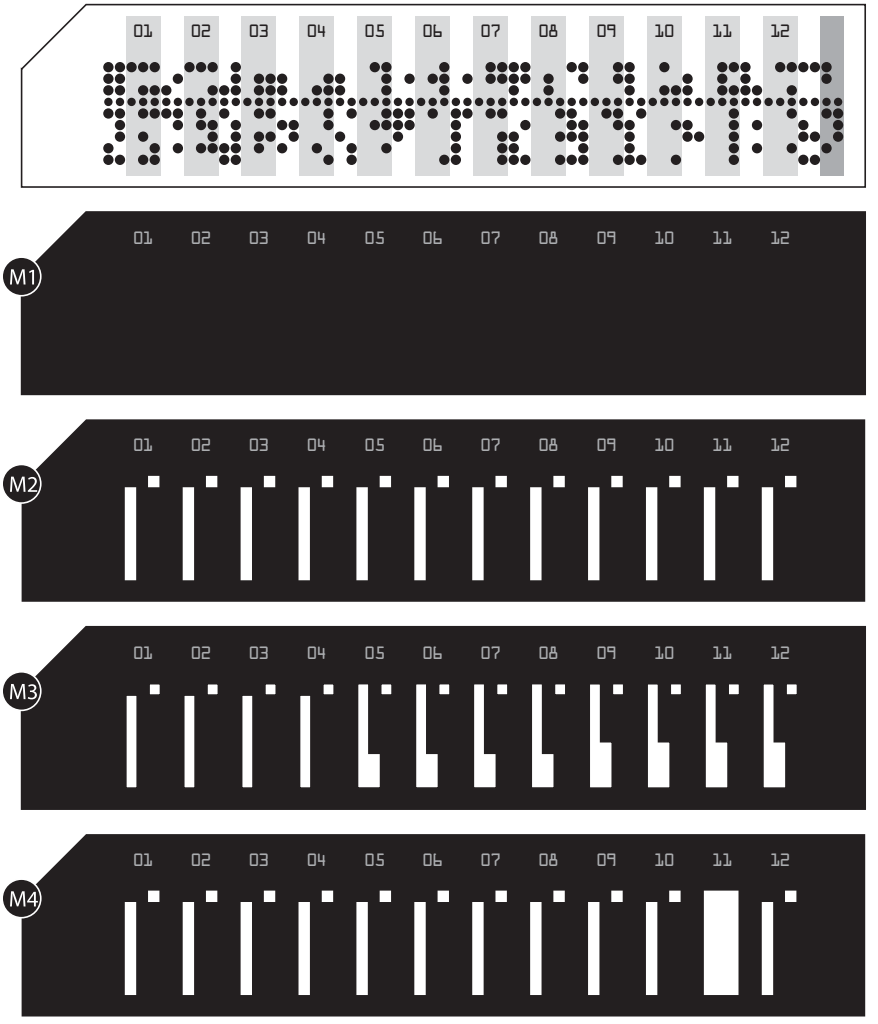

Figure 5. Secret template punched card (top) and a collection of possible masks that would be placed on top of the template card to cover sensitive information. Mask 1 covers the entire template, i.e., no information whatsoever is revealed. Mask 2 confirms that the counts are $<2^{18}$ in all bins; it also reveals the least significant bit for each bin. Mask 3 lowers the upper bounds, reproducing the situation shown in Figure 6; Mask 4 reveals Bin 11 entirely; this information could provide additional confidence in the calibration process, which relies on the 2.6-MeV peak that dominates the counts in that energy region.

particular steps that can be pursued independently. As is the case for other systems that permanently store the template, usually in encrypted form on an electronic storage medium, the reference item only needs to be available once, ideally, prior to inspection of the first candidate item. The advantage of the punched card/mask arrangement is that if the parties decide to make certain features of the template public later on, only the mask needs to be updated. The template and its cryptographic hash remain the same,

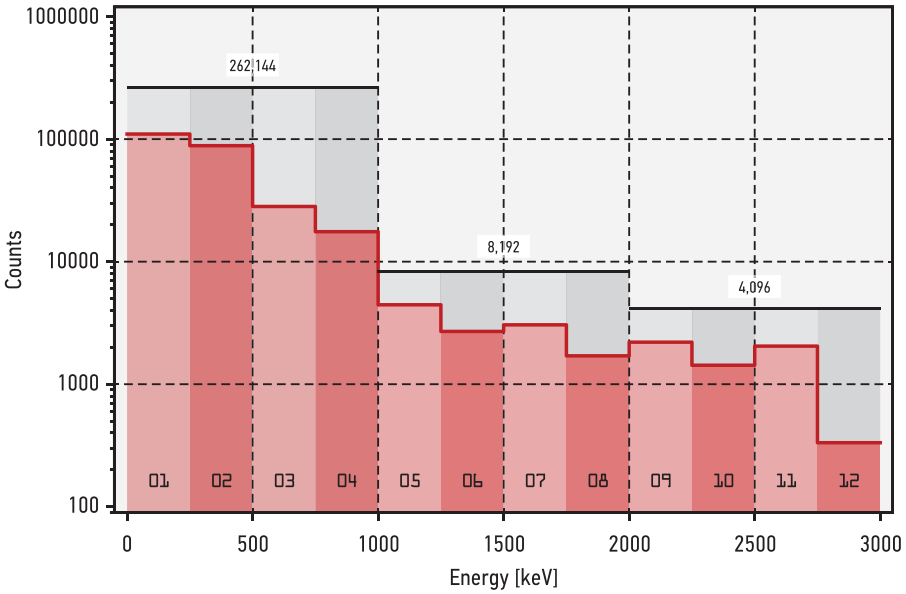

Figure 6. Making selected features of the template public. The host can choose to reveal upper bounds on the count limits for different energy bins. In this example, an appropriate mask can be used to confirm that the radiation signature of the inspected item results in fewer than 8,192 counts in Bins 05 through 08 and in fewer than 4,096 counts in Bins 09 through 12. Mask 3 in Figure 5 implements this particular scenario by revealing the relevant bits on the template card.

and no modifications to the system's software or mode of operation are necessary.

Figure 6 illustrates one example based on the signature of the 4.5-kg plutonium ball. The counts in most bins are orders of magnitude below the possible maximum, and the host party may find it unproblematic to reveal certain upper bounds for different energy bins. This process could also help confirm attributes of the inspected item such as bounds on the mass of plutonium or other relevant information about the warhead or parts being inspected.

Another possible advantage of revealing certain features of the template could be a calibration procedure that is more robust against possible deception efforts. For example, there may be concerns that the host could conceal radiation sources in proximity to the inspection equipment to throw off the calibration algorithm. Our system uses prominent peaks from the thorium decay chain (emitted by thoriated welding rods wrapped around the scintillator crystal) to calibrate the detector. The most important element of the procedure is the localization of the 2.614-MeV peak, which dominates the counts in Bin 11 (Figure 1). As a confidence-building measure, the parties could agree to reveal the counts in this bin entirely through a larger cutout in the mask (Figure 5, Mask 4). Since both parties know the

contribution of the calibration source, this procedure could confirm that the counts in Bin 11 are plausible.

Conclusion and next steps

This paper develops two new concepts for inspection systems based on the template-matching approach: First, it introduces physical templates using traditional punched cards to store secret data; second, it proposes the use of masks that can be placed on top of a punched card to selectively confirm particular features of the template. Operations emphasize manual procedures and visual confirmation and only require minimum dedicated hardware or software.

The idea of using physical templates seeks to address the challenge of establishing and maintaining confidence in the template, i.e., in the data characterizing the treaty accountable item. We believe that physical templates (for example, punched cards) may have important advantages compared to electronic templates stored in encrypted form on storage media that the host controls. We have used the SHA3-512 hash function to confirm that the template card has not been replaced or altered between inspections. The idea of revealing selected features of the template seeks to build additional confidence in the measurement process by constraining certain properties of an inspected item and by providing a mechanism to increase the level of transparency over time. Part of our broader concept of "vintage verification," we hope inspection systems based on physical templates could be more easily demonstrated than other systems and therefore contribute to efforts to support future reductions in the nuclear arsenals.

Acknowledgements

Alexander Glaser and Sébastien Philippe are grateful for the opportunity to conduct measurements at the Device Assembly Facility (DAF) made possible by the Consortium for Verification Technology under U.S. Department of Energy Award DE-NA 0002534.

Funding

NSF awards CCF 1565264 and CNS 1618026 and a Simons Fellowship supported the contributions by Boaz Barak.

Notes and References

1. K. D. Seager, R. L. Lucero, T. W. Laub, K. W. Inch, D. J. Mitchell, *Trusted Radiation Identification System (TRIS) User's Manual,* SAND2017-0578TR, Sandia National Laboratories, December 2002 (July 2011 Revision).

2. Defeating passive detection systems is not as straightforward as it might appear, however; see M. Göttsche, J. Schirm, and A. Glaser, "Low-resolution Gamma-ray Spectrometry for an Information Barrier Based on a Multi-criteria Template-Matching Approach," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 840 (2016): 139–144.
3. There exists one precedent for a U.S. information barrier being certified for use on a classified object in the presence of uncleared (Russian) visitors. The visitors, however, did not have the means to authenticate the information barrier. See J. Fuller, "Going to Zero: Verifying Nuclear Warhead Dismantlement," in *Cultivating Confidence: Verification, Monitoring, and Enforcement for a World Free of Nuclear Weapons,* ed. C. Hinderstein (Stanford: Hoover Press, 2010), 151.
4. Invented in the 18th century to control weaving looms, punched cards dominated the emerging electronic computer industry in the 1950s and 1960s. *The IBM Punched Card,* www-03.ibm.com/ibm/history/ibm100/us/en/icons/punchcard.
5. M. Kütt and A. Glaser, "Vintage Electronics for Trusted Radiation Measurements and Verified Dismantlement of Nuclear Weapons," *PLOS ONE* 14 (October 2019): e0224149.
6. The true measurement time could be shielded from the inspector by having the device wait until a certain (mutually agreed) time has passed before displaying any results.
7. M. Kütt, M. Göttsche, and A. Glaser, "Information Barrier Experimental: Toward a Trusted and Open-source Computing Platform for Nuclear Warhead Verification," *Measurement* 114 (2018): 185–190.
8. J. G. Fletcher, "An Arithmetic Checksum for Serial Transmissions," *IEEE Transactions on Communications* 30 (January 1982): 247–252.
9. Note that there may be other ways to confirm that the template card has not been modified, for example, using unique physical features; a punched card reader would still be required, however, to initialize the inspection system.
10. The hash shown here consists of 128 hexadecimal characters. Encoding one character requires 4 bits ($2^4 = 16$). Overall, there are $128 \times 4\,\text{bits} = 512\,\text{bits}$.
11. While this is a viable approach, it would be preferable (and also more elegant) if a process could be identified that relies exclusively on the physical properties of the card to confirm its authenticity. For example, such a process could leverage existing hardware security concepts such as physical unclonable functions or unique objects. See, R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld. "Physical One-Way Functions." *Science* 297 (2002): 2026–2030.
12. M. Kütt and A. Glaser, "Vintage Verification," *op. cit.*
13. In practice, the host and inspector would need to agree on the statistics they wish to use for the inspection; they also would have to authenticate the software running on the information barrier.
14. The template consists of 12 numbers, but the sum of these numbers is constant and known. In other words, if all but one values are known, then the last value of the template can also be determined. There are only 11 degrees of freedom.
15. M. Kütt, S. Philippe, B. Barak, A. Glaser, and R. J. Goldston, "Authenticating Nuclear Warheads With High Confidence," *55th Annual INMM Meeting,* Atlanta, Georgia, July 2014.
16. With regard to the sensitivity of this information, it is important to note that, under certain circumstances in the Unites States, gamma spectra of nuclear explosive devices are not considered classified as long as they are not associated with particular weapon

designs. For example, emergency responders can send radiation spectra "over the internet" to the NNSA Radiological Triage response team, which will send back an analysis of the data to the responders. See, *Handheld Radionuclide Identification Devices (RIDs),* Department of Homeland Security, "System Assessment and Validation for Emergency Responders (SAVER) TechNote," April 2014.