A PHYSICALLY CRYPTOGRAPHIC HOTELING OBSERVER FOR NUCLEAR WARHEAD VERIFICATION

by

Qing-Hua HE* *, Tian LI, Xiao-Suo HE, Kai-Kai LU, and Sheng-Kai WANG

College of Material Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Scientific paper
https://doi.org/10.2298/NTRP2104358H

When verifying the authenticity for nuclear warheads dismantlement toward a treaty partner's obligation, nuclear arms verification technologies are critical since only nuclear disarmament treaties are not sufficient to neutralize the existential threat of nuclear weapons. In this work, we present a verification method combining a numerical observer model and physical encryption techniques. The performance of the method is quantified by Monte Carlo simulations with several typical deception scenarios. Simulation results show this method can efficiently complete identification tasks in the presence of noise (< 5 %) and source-term variability, meanwhile exhibiting high security against brute-force attacks which reconstruct detection data by the exhaustive method.

Key words: non-proliferation, nuclear arms control, encrypted radiation imaging, mathematical observer model

INTRODUCTION

Nuclear arms control verification plays a key role in non-proliferation of nuclear weapons [1-3]. In the verification, inspectors need to make a confident conclusion on the authenticity of submitted items with very limited information about them, and no classified information is exposed to inspectors [4]. Radiation imaging technology can obtain detailed properties of the inspected object to evaluate the characteristics of nuclear weapons, but meanwhile, the imaging data might be used maliciously to reconstruct confidential information about the design/composition of the tested item [5]. To cope with the challenge of balancing the discriminability and security, verification technology of encryption from software and hardware have been developed [6, 7], for instance, the trusted radiation identification system and trusted radiation attribute demonstration system developed by Sandia National Laboratory, which use simple processors to isolate classified and unclassified information [8, 9]. However, complex analyses behind hardware also mean that inspectors have limited confidence in the results, and it may still leak sensitive information.

Alternative methods for reducing or modifying the general hardware encryption have been developed by numerous research groups. Glaser [10] has preloaded the detector based on the zero-knowledge protocol principle of information theory. If the aggregated output of preloaded data and object data reaches some predetermined value, it indicates that the test items are consistent with the declaration. But the existing zero-knowledge protocol methods have their complexity [10-12] and are still verification techniques that use hardware to protect sensitive information. It needs to set new preload data for the detector in the existence of source-term variabilities, such as source flux change or irradiation position change. Otherwise, it will output sensitive information to the inspectors. Gilbert and Jarman directly reduce the data dimensionality and cryptographically process the probe data by compressive sensing and hash function [13-15]. Although they belong to robust verification techniques, the system performance in the presence of source-item variability is still not discussed. This will become a new challenge in the application.

We introduce a gamma-ray-based verification method using a hardware encryption method and a dimensional reduction approach in the framework of the Hotelling observer (HO) model [16, 17], a mathematical observer model, which can maintain the same verification criteria when the source term changes. In this method, the hardware data encryption is realized via a random attenuating mask method. The HO model is used to reduce the dimension of the encrypted gamma-ray spatial information and to provide a statistical metric for judging whether the inspected item is a spoof or real. This method avoids generating high reso-

* Corresponding author; e-mail: heqh@nuaa.edu.cn

lution images or reconstructing spectral information on the object. Besides, the verification conclusion will not be affected by the change of source term thanks to the HO method. To test the performance of this method, numerical simulations are performed using the Geant4 toolkit [18]. We will show that a properly designed HO encoded imaging system can provide just such a measurement: the monitoring party can be allowed full access to the instrument before and after confirmation. Ideally, each detection event would be overwritten at the next detection, allowing sensitive data to be kept out of the system.

THEORY

The HO model is generally used in medical imaging to evaluate image quality and optimize the radiation dose [19-21]. It can simulate a human visual system to extract feature vectors and generate statistical decision metrics for defined tasks [22, 23]. Since the similarity between selecting the lesion area in the radiograph and warhead identification, we try to introduce the HO model into template-comparing-based warhead identification.

In this new method, a radiation measurement produces a unique data g_2 (where g is the energy deposition distribution after normalization and contains the 2-D imaging information) of an inspected item. Since g_2 contains sensitive information, we will delete them as soon as statistical metric are generated. The value of g_2 is compared against an authorized template's data g_1 to verify whether this inspected item is geometrically and materially identical with the template in terms of statistical metric λ

$$\lambda \quad W_H^T g_2 \tag{1}$$

where W_H is the Hotelling weight defined as,

$$W_H \quad K_g^{\ 1} \quad \bar{g} \tag{2}$$

$$K_g^{\ 1} \quad \frac{K_1 \quad K_2}{2} \tag{3}$$

$$\bar{g} \quad \overline{g_2} \quad \overline{g_1} \tag{4}$$

where $K_g^{\ 1}$ represents the average of two covariance matrices K_1 and K_2 of g_1 and g_2 respectively, and $\overline{g}$ – the difference between the average detection data g_1 and $\overline{g_2}$. Statistical decision metric λ – the projection of the original signal on the Hotelling weight space, and its distribution is an indicator of the possible changing of the inspected item. The values of λ will vary in a fixed region if the samples to be tested are consistent with the trusted template. Therefore, Hotelling-based verification provides an opportunity to distinguish fake objects against the Treaty Accountability Items (TAI) [9].

With this Hotelling decision metric, verification decisions can be made without reconstructing highly confidential geometric information on the inspected object. An inspector can only access the statistical metric, λ, and does not have access to the raw probing data. In addition to the data dimension reduction in the HO model, a physical encryption implementation is proposed to prevent possible confidential information disclosure. A lead plate containing randomly distributed holes is placed between the detector and the tested item to physically encrypt the profile of the transmitted photons which contains confidential information about the composition of the tested item. Then the HO is used to calculate the statistical metric of the warhead for the template matching task.

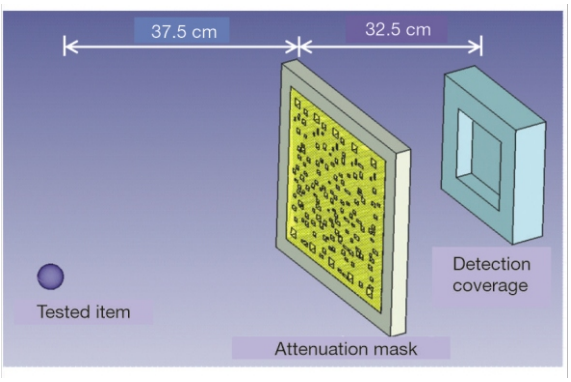

Figure 1. Schematic diagram of the HO encoded imaging system

NUMERICAL SIMULATION

This method is verified with Monte Carlo simulations using the Geant4 toolkit. In this section, we will demonstrate the implementation of our method and show that small differences between two inspected items can be reliably detected.

In the simulation, gamma rays of 1 MeV in energy generated uniformly in the x-y plane penetrate the tested sample. Transmitted rays are randomly attenuated and then recorded to construct the identity information of the tested sample, fig. 1. The attenuation information is obtained via recording the energy and flux of the gamma-rays passing through the tested sample and the randomly coded mask, that is placed between the tested sample and gamma-ray detector to encrypt the sensitive imaging data. A mask which is much larger than the size of the project to be tested is constructed by randomly digging holes in the lead plate. The position and size of the hole in the lead plate are defined by using the random Gaussian matrix. Randomly shifting the mask will result in random detection data, thus completing the encryption analysis of samples. To simplify the experimental scale, we only simulate the effective area of the mask which is on the same horizontal line as the sample to be tested. The HO model is used to generate a statistical metric

Table 1. Spoof items

Spoof	Size [cm]	Material
1	2.3 cm	Highly enriched uranium
2	2.4 cm	Highly enriched uranium
3	2.45 cm	Highly enriched uranium
4	2.49 cm	Highly enriched uranium

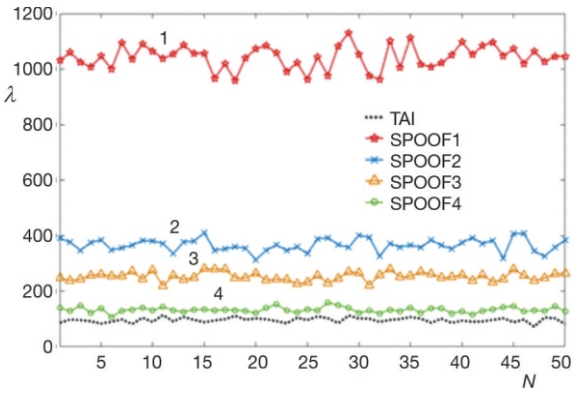

Figure 2. The statistic metric values of the true item and four spoof items

with the incoming signal of the randomly attenuated gamma-rays.

To test the capability of identifying possible deception against the true sample, we take a solid uranium sphere of 2.5 cm in radius as the TAI, and a set of balls with a radius less than 2.5 cm as the spoofs, placed between a strong gamma-ray source and a position-sensitive detector. The chosen spoof items are summarized in tab. 1.

For each sample, two simulations are performed. The first simulation is used for obtaining the statistical metric, λ, of the HO-based inspection concept, and the second is used to test the accuracy of the verification method. The statistic metric, λ, of the tested sample is used for authenticity judgment by comparing it with a set value. Generally, there exists a distinct boundary zone between the true and spoof samples' statistic metric values, as shown in fig. 2 in which the statistic metric, λ, of the true reference sample and four spoof samples is plotted as the attenuating mask has existed. For the true sample, the statistical metric, λ, varies in the region of $81 < \lambda < 111$, while the values of λ for the spoofs are outside this window.

The robustness of the Hotelling examination instrument requires that the values of λ for identical items should be close to each other in the presence of extraneous interference conditions in the system (*e. g.*, noise). To quantify the performance of the Hotelling examination instrument, the gamma-ray imaging data of the spoof items and TAI are obtained with the same encoded shielding mask. The variation of imaging data is simulated by adding white noise with normal distribution to the signal, varying from 0.5 % to 10 %. The noise signal depends on the signal energy

$$P_n \quad P_s \quad \varepsilon \tag{5}$$

where P_s is the signal power obtained from the probe data, and P_n – the noise power, and ε – the percentage of noise.

The threshold of decision statistics is adjusted to draw the receiver operating characteristic (ROC) curve, which describes the relationship between sensitivity and specificity [24]. The ROC curve reflects the model's sensing of external stimuli and can be used to assess the classification ability under different disturbances. When the ROC curve is above the diagonal and deviates from the diagonal, the stronger the discrimination of the analysis method is. To further evaluate the system robustness, we changed the source flux in simulation and used the area under the ROC curve – AUC, to describe its performance when combined with the changes in the percentage of noise.

The security of the Hotelling verification means that the sensitive information cannot be determined by those with access to the output and that the physical encryption performance depends on the exact location of the object under test relative to the mask. Actually, the inspectors are unable to reconstruct data relying only on the statistical metric, but we still do not exclude the possibility of violent decryption of the detection process by brute force based on the exhaustive method. We try to verify all possible cases one by one until all cases are verified and quantify the security of the Hotelling verification technique by reconstruction accuracy. In our prediction, the time consumed to decipher the original image using known information will be huge and the reconstruction accuracy is extremely low.

RESULTS

The ROC curve is used to quantify the ability of the HO model to correctly classify samples and serve as an indicator of system robustness. 1000 project decision statistics and real project decision statistics are generated, and a confidence threshold is set to declare images with a higher than that range as a spoof. By varying the threshold, ROC curves are produced, certificating the accuracy and sensitivity of the Hotelling-based verification, as shown in fig. 3. The curves at the noise levels of 0.5 % and 1 % almost have the best performance and hence overlap each other. In this example, the ROC curve shows the strong resolution of Hotelling analysis technology. Setting λ to 81-111 will get 99 % resolution accuracy.

The study of the noise influence on the robustness is limited to the different percentages of white noise. At the noise level of 5 %, the performance is still ideal, and the area value under the curve is about 0.9. The robustness of the system gets worse rapidly with the increase of noise. The optimal threshold can be found in the ROC curves by finding the closest point to the left-upper corner. In our work, a threshold of 115 is chosen, leading to minimum errors in discriminating between TAI and spoofs.

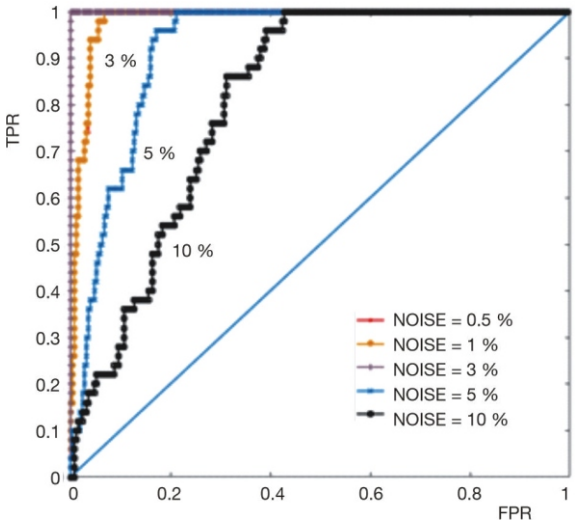

Figure 3. The ROC curves at different percentage noise

To investigate the resolution of the Hotelling observer in the presence of source variability, the gamma-ray energy, and the counts are varied in the simulation as listed in tab. 2. With these changes and noise level changes, tab. 3, areas under these ROC curves, AUC, are shown in fig. 4. The threshold value of 0.8 shown in the purple plane is an indication of the *good* performance of the measurement system. Figure 4 shows that the area under the ROC curve, AUC, is higher than this threshold. If needed, the system's robustness can be improved by reducing noise.

In this method, security is ensured by the attenuation mask and data dimensionality reduction. Therefore, the physical security depends on the position of the item to be measured relative to the mask. The system can only share the basic information of mask initial design and decision variable value with inspectors. The imaging information is removed immediately after the model generates decision statistics in electronic instruments. With these secure procedures, the inspectors cannot reconstruct the sensitive information of nuclear warheads only by using decision statistics. Figure 5

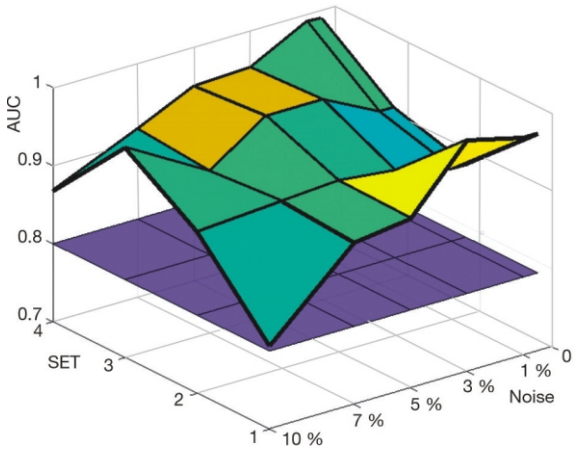

Figure 4. The AUC values when noise levels and source changes

Table 2. Inspection sets

Set	TAI source		Inspected item source	
	Type	Counts	Type	Counts
1	1 MeV gamma-ray	3000000	1 MeV gamma-ray	3000000
2	1 MeV gamma-ray	3000000	1 MeV gamma-ray	1000000
3	1 MeV gamma-ray	3000000	1 MeV gamma-ray	5000000
4	1 MeV gamma-ray	3000000	0.8 MeV gamma-ray	3000000

Table 3. The AUC value

Noise set	0.5 %	1 %	3 %	5 %	7 %	10 %
1	0.978	0.978	0.996	0.917	0.907	0.806
2	0.893	0.894	0.911	0.919	0.914	0.907
3	0.919	0.919	0.956	0.956	0.906	0.969
4	0.988	0.989	0.951	0.948	0.915	0.867

shows the encrypted imaging information of 50 detection samples, each of which is represented by a row obtained by summing each column of the original detected 50 50 matrix and normalizing the summation value as $SUM_i = (SUM_i - MIN)/(MAX - MIN)$ ($i = 1, 2, ..., 50$), where MIN and MAX represent the minimum and maximum values among the 50 summation values, resprectively.

Here we simplify the exhaustive attack only for extreme information leakage scenarios. The reconstruction accuracy of the exhaustive attack is quantified as the structural similarity, r, between the reconstructed image, A, and the real sensitive image, B.

$$r \quad \frac{\sum_M \sum_N (A_{MN} - \overline{A})(B_{MN} - \overline{B})}{\sqrt{\sum_M \sum_N (A_{MN} - \overline{A})^2 \sum_M \sum_N (B_{MN} - \overline{B})^2}} \quad (6)$$

where $\overline{A}$, $\overline{B}$ are the pixel averages of the simulated reconstructed image and the detected image, A_{MN}, B_{MN} are the elements of the M^{th} row and N^{th} column of the image matrix, and, r, close to 1 indicates that the two

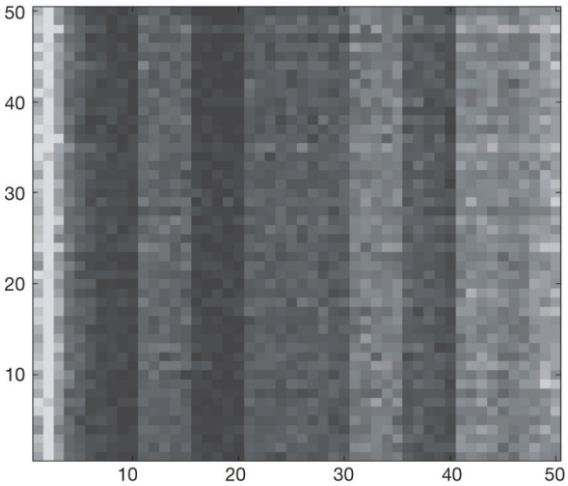

Figure 5. Enerypted imaging information of multiple detection samples

images are strongly correlated. The maximum reconstruction accuracy does not exceed 0.5 when the inspector has full access to the λ through 80 % of the effective area of the mask. Although more effective or drastic methods may exist in the future for brute force cracking of sensitive information, the attempt for the computation time of the exhaustive method is sufficient to show the high security of the Hotelling-based verification. In addition to this, the higher complexity of the actual verification object compared to the simulated object described in this paper will further increase the complexity of the method.

CONCLUSION

Using the linear HO model combining a randomly attenuating mask, the nuclear arms verification method we propose shows its ability to identify false inspected items from the true template. Numerical simulations show this method increases the difficulty of reconstructing sensitive data of inspected weapons. In addition, to measure the robustness of the inspection system the ROC curves is employed to analyze its performance at different white noise level and different radiation source setting conditions. Monte Carlo simulations show this verification system still works well at a noise level of 5 %.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China, Grant No. 11805099, and the Fundamental Research Funds for the Central Universities, Grant Nos. NS2018043, NJ2020017-5, and the Foundation of Jiangsu Province high-level innovation and entrepreneurship talent introduction plan, grant No. 1006-KFR20052

AUTHORS' CONTRIBUTIONS

Q. H. He and T. Li proposed the idea of this nuclear warhead verification system with a physically cryptographic Hoteling observer method. A numerical simulation was performed by T. Li and X. S. He. The manuscript was compiled by T. Li, Q. H. He, and X. S. He. The literature search was performed by K. K. Lu, S. K. Wang. All authors participated in data analysis and discussion under the supervision and guidelines of Q. H. He.

REFERENCES

[1] Rodriguez, R., *et al*., A Hybrid Model to Deal with Missing Values in Nuclear Safeguards' Evaluation, *International Journal of Nuclear Knowledge Management, 5* (2011), 2, pp. 194-218

[2] Christoff, J. A., State Department: Key Transformation Practices Could Have Helped in Restructuring Arms Control and Nonproliferation Bureaus, *Government Accountability Office Reports, 71* (2009), 1, pp. 319-330

[3] Holdren, J. P., Fetter, S., Monitoring Nuclear Weapons and Nuclear-Explosive Materials: An Assessment of Methods and Capabilities, 2005

[4] Hinderstein, C., Cultivating Confidence, *Journal of Nuclear Materials Management, 39* (2011), 2, p. 33

[5] Abe, M. C., Reentry Vehicle On-Site Inspection Technology Study, Technical report, 6 March 1992-19 May 1993, 1994

[6] Lancor Deceased, J. H., Rak, A., Information Carrier, ed, 1975

[7] Brubaker, E., *et al*., Information Barriers for Imaging (Conference: Proposed for Presentation at the WMS Review Meeting Held March 17-19, 2015 in Livermore, CA.). ; Sandia National Lab. (SNL-CA), Livermore, CA (United States), 2015, p. Medium: ED; Size: 1 p.

[8] Mitchell, D. J., Tolk, K. M., Trusted Radiation Attribute Demonstration System (Conference: Institute of Nuclear Materials Management 41st Annual Meeting, New Orleans, LA (US), 07/16/2000--07/20/2000; Other Information: PBD: 13 Jun 2000). ; Sandia National Labs., Albuquerque, NM (US); Sandia National Labs., Livermore, CA (US), 2000, p. Medium: P; Size: 8 pages

[9] Flynn, *et al*., Next Generation Trusted Radiation Identification System (NG-TRIS), 2010

[10] Glaser, A., *et al*., A Zero-Knowledge Protocol for nuclear Warhead Verification, *Nature, 510* (2014), 7506, pp. 497-502

[11] Brennan, J., *et al*., Demonstration of Two-Dimensional Time-Encoded Imaging of Fast Neutrons, *Nuclear Inst & Methods in Physics Research A, 802* (2015), Dec., pp. 76-81

[12] Kemp, R. S., *et al*., Physical Cryptographic Verification of Nuclear Warheads, *Proceeding*, *National Academy of Sciences of the United States of America, 113* (2016), 31, pp. 8618-8623

[13] Zhen-Kun, W., *et al*., A Robust and Discriminative Image Perceptual Hash Algorithm, in Fourth International Conference on Genetic & Evolutionary Computing, 2010

[14] Jarman, K. D., *et al*., Low-Intrusion Techniques and Sensitive Information Management for Warhead Counting and Verification: FY2011 Annual Report. Office of Scientific & Technical Information Technical Reports, 2011

[15] Gilbert, A. J., *et al*., A Single-Pixel X-Ray Imager Concept and Its Application to Secure Radiographic Inspections, *Nuclear Instruments & Methods in Physics Research, 861* (2017), July, pp. 90-97

[16] Popescu, L. M., Lewitt, R. M., Comparison Between TOF and Non-TOF PET Using a Scan Statistic Numerical Observer, in Nuclear Science Symposium Conference Record, 2007

[17] Brankov, J. G., *et al*., Learning a Nonlinear Channelized Observer for Image Quality Assessment, In 2003 IEEE Nuclear Science Symposium, Conference Record (IEEE Cat. No. 03CH37515) (Vol. 4, pp. 2526-2529). IEEE.

[18] Agostinelli., S., Geant4: A simulation Toolkit, *Nuclear Instrumentation and Methods in Physics Research A, 506* (2003), 3, pp. 250-303

[19] Jr, F. A. M., *et al*., Radiologic and Nuclear Medicine Studies in the United States and Worldwide: Frequency, Radiation Dose, and Comparison with Other

Radiation Sources--1950-2007, *Radiology, 253* (2009), 2, pp. 520-531
[20] Brenner, D. J., Hall, E. J., Computed Tomography--An Increasing Source of Radiation Exposure, (in eng), *N Engl J Med, 357* (2007), 22, pp. 2277-2284
[21] Barrett, H. H., *et al*., Task-Based Measures of Image Quality and Their Relation to Radiation Dose and Patient Risk, *Physics in Medicine & Biology, 60* (2015), 2, pp. 1-75
[22] Wunderlich, A., Noo, F., Image Covariance and Lesion Detectability in Direct Fan-Beam X-Ray Computed Tomography, *Physics in Medicine & Biology, 53* (2008), 10, p. 2471
[23] Leng, S., *et al*., Correlation Between Model Observer and Human Observer Performance in CT Imaging When Lesion Location is Uncertain, *Medical Physics, 40* (2013), 8
[24] Wunderlich, A., Noo, F., Estimation of Channelized Hotelling Observer Performance With Known Class Means or Known Difference of Class Means, *IEEE Transactions on Medical Imaging, 28* (2009), 8, p. 1198

Received on October 31, 2021
Accepted on February 8, 2022

Ћинг-Хуа ХЕ, Тјен ЛИ, Сјао-Суо ХЕ, Кај-Кај ЛУ, Шенг-Кај ВАНГ

КРИПТОГРАФСКО ФИЗИЧКИ HOTELLING ОСМАТРАЧ ЗА ВЕРИФИКАЦИЈУ НУКЛЕАРНЕ БОЈЕВЕ ГЛАВЕ

Приликом провере аутентичности демонтирање нуклеарних бојевих глава у складу са обавезом партнера из споразума, технологије верификације нуклеарног оружја су критичне јер споразуми о нуклеарном разоружању нису сами довољни да неутралишу егзистенцијалну претњу нуклеарног оружја. У овом раду представљамо методу верификације која комбинује нумерички модел осматрања и технике физичког шифровања. Перформансе методе су квантификоване Монте Карло симулацијама са неколико типичних сценарија преваре. Резултати симулације показују да ова метода може ефикасно да обави задатке идентификације у присуству шума (< 5 %) и варијабилности извора, при томе показујући високу сигурност од напада грубом силом који реконструишу податке детекције исцрпним поступком.

Кључне речи: контрола нуклеарног оружја, шифровано радијационо снимање, математички модел осматрача