Optical DNA

Deepak Vijaywargi[†], Dave Lewis[‡], and Darko Kirovski[◇]

[†] Dept. of Electrical Engineering, University of Washington, Seattle, WA 98195, USA
[‡] Microsoft Corp., One Microsoft Way, Redmond, WA 98052, USA
[◇] Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA

Abstract. A certificate of authenticity (COA) is an inexpensive physical object with a random and unique structure S which is hard to near-exactly replicate. An inexpensive device should be able to scan object's physical "fingerprint," a set of features that represents S. In this paper, we explore one set of requirements that optical media such as DVDs should satisfy, to be considered as COAs. As manufacturing of such media produces inevitable errors, we use the locations and count of these errors as a "fingerprint" for each optical disc: its optical DNA. The "fingerprint" is signed using publisher's private-key and the resulting signature is stored onto the optical medium using a post-production process. Standard DVD players with altered firmware that includes publisher's public-key, should be able to verify the authenticity of DVDs protected with optical DNA. Our key finding is that for the proposed protocol, only DVDs with exceptional wear-and-tear characteristics would result in an inexpensive and viable anti-counterfeiting technology.

1 Introduction

Counterfeiting is regarded as a form of illegal trade where the seller fools the buyer into believing that the merchandise is authentic and collects the full "legal-market" price on the product. The counterfeiter usually earns profit margins that are higher than that of the original manufacturer due to lack of development and marketing costs. The software industry has suffered from this problem since the inception. To date, a few tools have been efficient in attenuating counterfeiting.

Since the early work out of Sandia National Labs by Bauder and Simmons [1], certificates of authenticity have attracted attention as a possible remedy. A *certificate of authenticity* (COA) is a digitally signed physical object with a random unique structure such that: **R1** – the cost of creating and signing original COAs is small, **R2** – the cost of manufacturing a COA instance is substantially lower than the cost of its near-exact replication, **R3** – the cost of verifying the authenticity of a signed COA is small, and **R4** – a COA must be robust to ordinary wear and tear. In essence, COAs connect the physical and digital world into a unifying concept that could be applied to a variety of security applications, ranging from anti-skimming for credit cards to tamper-evident seals [2].

In this paper, we propose COAs built based upon the fact that optical media, even when freshly imprinted, still have numerous errors due to the nature of their manufacturing process. There are four sets of detectable errors that occur: (**e1**) – for each disc imprinted using the same "negative," (**e2**) – uniquely per disc and their detection is nearly deterministic, (**e3**) – uniquely per disc however the likelihood that they are detected is ∼0.5, and (**e4**) – due to wear and tear. Sets **e1**–**3** occur at manufacturing, while the set **e4** increases throughout the lifetime of the disc. Although production errors can be controlled as the adversary would stamp discs using a "negative" that already has desired errors imprinted, the adversary cannot control the rate of additional inevitable errors (**e1**–**3**) using a low-cost manufacturing process and materials. Thus, the expectation is that "counterfeit" DVDs would always have at least twice as many errors as "authentic" ones for comparable printing technologies. We denote "fingerprints" constructed based upon such errors: *optical DNA* (o-DNA). For widely accepted o-DNA, costs related to **R1** and **R3** would be negligible.

Using o-DNA within the cryptographic realm is simple. When creating an o-DNA instance, i.e., optical disc, the publisher digitally signs the positions of manufacturing errors on this disc using a traditional PKCS [3] as follows. First, the "fingerprint" is scanned using a standard DVD player modified to output the low-level errors, then compressed into a fixed-length string f. Arbitrary text t associated with the disc is then concatenated to f, $w = f||t$, hashed, and signed using the private key of the issuer. Next, the resulting signature s, w, and optionally, publisher's certificate, are encoded onto the o-DNA instance using a post-production mechanism. SONY, for example, offers a technology that allows for several hundred bytes to be imprinted onto a disc post molding and bonding.[1]

Verification of o-DNA instances is straightforward provided that the verifier is in possession of publisher's public key. We assume that a malicious party cannot tamper with a specific verifier used in-field, however, verifier's full design spec is considered public knowledge. The verifier does not need to store any secrets to fulfill its basic task. The key component of the verifier is a function, $d(f, f')$, that computes the proximity of the signed and in-field scanned "fingerprint." If s is valid and $d(f, f') < \delta$ then the instance would be deemed authentic (δ is a relatively small constant). The system should tolerate a relatively high rate of false negatives because publishers can choose to react only if they receive uncharacteristically high ratio of "false negatives" from a specific source. Figure 1 details the issuing and verification of o-DNA instances using a block diagram.

Finally, we refer the Reader to review a short survey of related work on COAs in [2]. We also mention that there exist several technologies for copy protection of DVDs[2], all with a common problem: all bits that contribute to the protection and are readable by a standard DVD player are easy to circumvent in software. A rare instance of relative success is Microsoft's XBOX which uses a distinct obfuscated low-level data/track format substantially different from DVD-R and a custom DVD player that can read such optical discs.

[1] Unfortunately, we are not aware of any technical references to this technology, hence we refer to it as personal communication with SONY.

[2] See informal survey at Wikipedia: `http://en.wikipedia.org/wiki/CD_copying_software`.

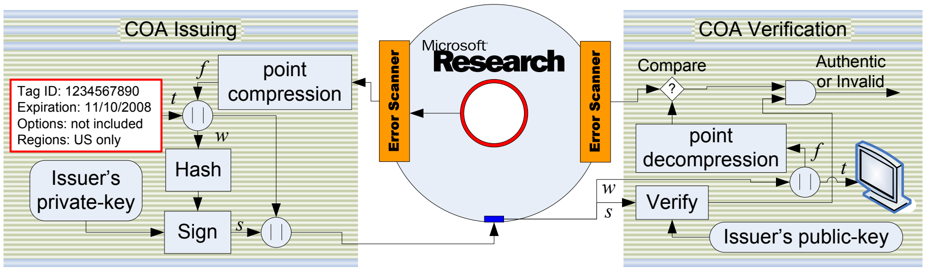

Fig. 1. Diagram of actions taken while signing and verifying an o-DNA instance.

2 Optical DNA

Here, we describe how o-DNA is constructed and review its security features. The 120mm DVD-R standard is detailed in [4]. Impression-based manufacturing of DVD-Rs is a well understood process with low variance of output produced within the same manufacturing facility; however, with possibly strong variance of output across different facilities – in particular for low-quality manufacturing.

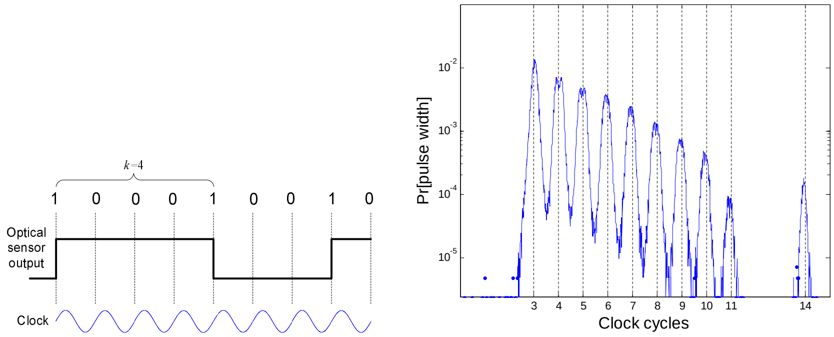

Fig. 2. (left) An example of encoding 100010010 using an NRZI encoder. (right) Distribution of pulse-widths for t_i over the 24^{th} millimeter of a high-quality DVD with the installation data for Microsoft Visual Studio 5.0.

The sensor readout of the physical specification from a DVD-R consists of an NRZI-encoded signal clocked at 26.1MHz [4]. The signal is "high" or "low" depending on whether there is a pit or a land on the optical disc. The NRZI encoding is such that between two "ones" (i.e., signal floor changes) the signal stays at the same level for integer $k \in \mathbb{C}, \mathbb{C} \equiv \{3, 4, 5, 6, 7, 8, 9, 10, 11, 14\}$ number of clock cycles. The encoding is illustrated in Figure 2(left). Due to manufacturing inefficiencies, in general the distance between two signal floor changes is not an exact multiple of the master clock cycle – it is rather a random variable t that can be represented as: $t_i \equiv k_i + \mathcal{N}(0, \sigma_M), k_i \in \mathbb{C}$, where $\mathcal{N}(0, \sigma_M)$ denotes a random zero-mean Gaussian variable with standard deviation equal to σ_M.

Generally we recognize that high-quality manufacturing should have relatively low σ_M. We assume in o-DNA that the legal publisher of protected DVDs is using state-of-the-art manufacturing, i.e., that it is hard to achieve significantly better error rates by an adversarial manufacturing process. Although the error model is likely to be smooth over $|t_i - k_i|$, for a small ε, we postulate:

(i) Probability that a signal with $\frac{1}{2} - \varepsilon < |t_i - k_i| < \frac{1}{2} + \varepsilon$ is incorrect, is 0.5.
(ii) Probability that a signal with $\frac{1}{2} - \varepsilon \geq |t_i - k_i|$ is incorrect, is 0.
(iii) Probability that a signal with $\frac{1}{2} + \varepsilon \leq |t_i - k_i|$ is incorrect, is 1.

Figure 2(right) presents the distribution of pulse-widths t_i over the 24^{th} millimeter of a single high-quality DVD with the installation data for Microsoft Visual Studio 5.0. The probability that t_i is close to an integer value is relatively high and conversely the probability that t_i is half-way between two integers, is around two orders of magnitude lower. To estimate the error rate, in Figure 3(left) we plot the distribution of $\varepsilon = ||t_i - k_i| - 0.5|$ over the same disc instance used to plot the distribution in Figure 2(right). The data was collected using a reference DVD player by AudioDev with an analog TTL-output representing the NRZI encoded signal recorded at the output of the optical sensor in the DVD player [5]. The TTL-output was sampled at a rate of 10Gsamples/sec to produce accurate statistics. Figure 3(left) illustrates that the likely error rate on the disc used in the experiment, assuming an error threshold $\varepsilon \in [0.05, 0.1]$ and that $\Pr[\frac{1}{2} + \varepsilon \leq |t_i - k|] = 0$, is roughly on the order of 10^{-3}.

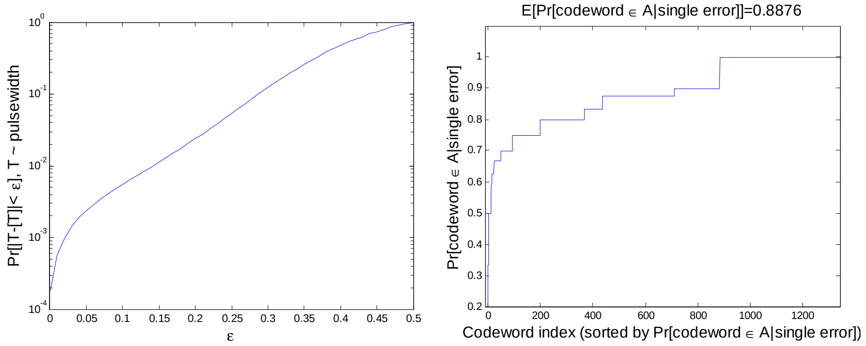

Fig. 3. (left) Distribution of $\varepsilon = ||t_i - k_i| - 0.5|$ over the 24^{th} mm of a DVD with the installation data for MS Visual Studio 5.0. (right) Probability of illegal symbol after an occurrence of a single-position error on a legal 16-bit symbol from $\mathcal{A}$. Symbols from $\mathcal{A}$ are sorted based upon the resulting probability, i.e., ∼30% of all symbols in $\mathcal{A}$ never produce an error detectable during NRZI decoding.

The DVD-R standard uses an efficient codec for converting an alphabet $\mathcal{A}$ that consists of 16-bit symbols encoded using NRZI, into an alphabet $\mathcal{L}$ of 256 8-bit words [4]. Not all 16-bit symbols belong to $\mathcal{A}$, hence we distinguish between legal (that belong to $\mathcal{A}$) and illegal 16-bit symbols. Figure 3(right) illustrates the probability that a legal 16-bit keyword remains legal after the event of an arbitrary single position error. Since the probability of an error is relatively low, we consider only the case when a symbol from $\mathcal{A}$ is affected by only one error. The overall probability that a 16-bit erroneous symbol cannot be found in the look-up table $\mathcal{A} \rightarrow \mathcal{L}$, is roughly $p = 10^{-1}$. That means that although there exists an error on the optical disc, the likelihood p that it will be detected during NRZI decoding is low. Such errors are detected accurately in higher levels of decoding.

The main synchronization primitive for low-level encoding in the DVD-R standard is a cluster of 26 data fields. Each field consists of a specific synchronization pattern (32 NRZI-bits long) and a payload of 91 symbols from $\mathcal{A}$ (1456 NRZI-bits payload). The synchronization pattern is a 32-bit synchronization symbol selected from a specific 32-symbol alphabet $\mathcal{S}$ [4]. The 38688-bit clusters represent the main storage unit on a DVD-R. We classify all error cases as:

a) **illegal codeword** – (32%) a payload symbol is altered due to an error; the resulting codeword cannot be found in the set of legal words $\mathcal{A}$.
b) **codeword still in $\mathcal{A}$ after error** – (not detected) a payload symbol is altered due to an error; the new symbol exists in $\mathcal{A}$.
c) **shift required to correct a synchronizing symbol** – (63%) errors commonly shift the synch symbols with respect to their correct position within a cluster. Typically, adjustment shifts for one or two positions are sufficient to realign the synch symbols.
d) **illegal synch codeword** – ($< 1\%$) a synch symbol is altered due to an error; the new codeword is not found in the set of legal synch codewords $\mathcal{S}$.
e) **all zeroes codeword** – (4%) – all bits of a symbol equal zero. Such a symbol is not legal both in $\mathcal{A}$ and $\mathcal{S}$; it deserves special attention because it corresponds to a specific manufacturing error.

Percentages presented immediately after the item title in the previous list, specify the occurrence rate for each error type that we detected on the 24^{th} millimeter of our DVD disc under test. Since we did not soft-decode data past the EFM decoding step, we were not able to identify errors of type **b)**. It is expected that the number of such errors is ~10x greater than errors of type **a)**.

Since many manufacturing errors manifest as signals with pulsewidths far from integer clock values, some of these errors will be read differently during distinct DVD read-outs. For example, assume a pulsewidth $d_i = 3.501$ clock cycles. A DVD player could read this pulsewidth as 2 or 3 zeroes in different read-outs. Clearly only one of the values is correct, whereas the other one is erroneous. Since this is a probabilistic effort, while both issuing and verifying the errors of the o-DNA, the player needs to read the same track several times in order to detect most errors. Based upon our error model, reading L times the desired set of tracks from the DVD that contain the "fingerprint," would be sufficient to detect at least $1 - 2^{-L}$ of **e3** errors in that region.

Verification of the o-DNA consists of two steps:

I **verifying that the in-field disc is the same as the issued one** – when scanning, it is trivial for the publisher to identify errors of type **e1** – by comparison with media printed from the same "negative," **e2** and **e3** – by the likelihood of detection in multiple readouts. Dr. Holger Hoffman from Technicolor Inc. has estimated that for a specific sample of DVDs manufactured at their facilities, the ratio of error sets **e1**:**e2**:**e3** is 65:17:28. The o-DNA issuer would sign all of them including their types. During verification, the in-field multi-scan (L times) of errors should identify all errors of type **e2** and most errors of type **e1** and **e3**. Thus, we use the following detector in this step: $||\mathbf{e1} \cap \mathbf{et}|| \geq \alpha_{\mathbf{1}}||\mathbf{e1}||$, $||\mathbf{e2} \cap \mathbf{et}|| \geq \alpha_{\mathbf{2}}||\mathbf{e2}||$, and $||\mathbf{e3} \cap \mathbf{et}|| \geq \alpha_{\mathbf{3}}||\mathbf{e3}||$, where constants $\alpha_1 = \alpha_2 \approx 1$ and α_3 is relatively close to 1 but proportional to L. Operator $||\cdot||$ returns the cardinality of the argument. Set **et** represents all the errors extracted during an in-field test of an o-DNA instance.

II **verifying that the in-field test does not yield too many errors**; the adversary can imprint error sets **e1**, **e2**, and **e3** during an adversarial effort and thus, create a match in step I. However, she cannot control the manufacturing process to the extent to prevent additional expected manufacturing errors. Therefore, the expectation is that she will produce approx. $||\mathbf{e1} \cup \mathbf{e2} \cup \mathbf{e3}||$ additional errors on the counterfeit disc using a printing technology similar to the publisher's. Therefore, the verifier must check whether $||\mathbf{et}|| \leq ||\mathbf{e1} \cup \mathbf{e2} \cup \mathbf{e3}||(\mathbf{1} + \beta)$, where β is a real positive scalar smaller than but relatively close to 1 (e.g., $\beta = 0.8$).

Assuming that there is no adversarial attack, the probability of a false positive is practically equal to zero even for relatively small $||\mathbf{e1} \cup \mathbf{e2} \cup \mathbf{e3}||$. The probability of a false negative is proportional to the α parameters and can be tuned to be relatively low. It is rather important that the cardinality of the set of additional errors due to wear and tear $||\mathbf{e4}||$ is not greater than $\beta||\mathbf{e1} \cup \mathbf{e2} \cup \mathbf{e3}||$ – in the opposite case, the verifier would report false negatives. This is a crucial issue with the proposed technology as current wear and tear characteristics of DVDs are far from acceptable [6]. Thus, our key conclusion is that o-DNA, as defined, would be applicable only to DVDs with superior wear and tear characteristics – clearly, scratch resistant materials and more sophisticated sensors would have to be used to enable o-DNA. Another critical comment is the fact that algorithms for symbol decoding are not mandated by the ECMA standard – thus, different manufacturers may use different multiword, usually Viterbi, decoders that could impact error detection. To enable o-DNA, the word decoders in a DVD player would need to be standardized. However, once the standardization is established, o-DNA would represent an exceptionally inexpensive way to identify authentic DVDs, a tool that could be essential in fighting counterfeiting. As expected, the converse part of the grey market where the buyer willingly purchases an obviously pirated DVD copy cannot be addressed by any anti-counterfeiting technology.

Finally, we consider an implementation of o-DNA, where at an error rate of 10^{-3}, an error read-out from the 24th millimeter (approx. 10^3 revolutions) of a standard DVD-R disc, is sufficient to produce $||\mathbf{e1} \cup \mathbf{e2} \cup \mathbf{e3}||$ on the order of 10^2. The resulting o-DNA message stored back onto the DVD would be approx. 1Kb long. Since the disc encounters 24 revolutions per second at 1x playback speed, one can observe that the verification of an o-DNA could be done in approx. L seconds at 32x playback speed.

3 The DVD-R Manufacturing Process

DVD-R media is created using a high-speed automated replication process. Initial glass master of data to be used for disc creation is created via a photolithography process using a laser beam recorder to expose a photo resist coated blank glass master. For DVD5 a single glass master is required as data is wholly contained on one layer of the disc. The glass master is "developed" after exposure resulting in a pattern of bumps in the remaining photo resist. The glass master is nickel metal plated to create a "father," a mirror image negative of the data created by the laser beam recording process on the glass master. The "father" is separated from the glass master and plated with nickel again to create a "mother" positive (same as the original glass master). One "father" can create 5 to 20 "mothers." Each "mother" is again nickel plated to create a stamper; a single "mother" can create up to 50 stampers, the stamper is again a negative image of the original data created by the laser beam glass mastering, each stamper can create up to 10^5 discs. The stampers are separated from the "mother" after plating, and then "punched" to correct outside diameter and correct hub hole inner diameter as required for the specific molding equipment.

The punched stamper is mounted inside the molding chamber of the manufacturing line. Molten polycarbonate is injected under pressure, heat and humidity into the mold chamber. The pattern of pits and lands on the stamper are impressed into the clear polycarbonate under several tons of pressure. The polycarbonate is rapidly cooled via chilled water flowing through the mold chamber housing and separated from the stamper and ejected from the mold chamber. This is considered a DVD half disc, as it is one layer of the final DVD. At this point the disc would not reflect a laser beam in the DVD player. For DVD5 the following steps are then executed. The ejected clear polycarbonate is plated with a layer of reflective material such as aluminum using a sputtering process in order to reflect the laser beam in the DVD player. A clear half disc is bonded onto the aluminum coated half disc creating a final disc 1.2mm thick, with the data in the middle of the disc at ∼0.6mm from the bottom surface.

In each of the steps above mechanical tolerances will be present. The degree of jitter and degree of run out in the original glass master will set a baseline for the final finished discs as to the number of errors present. As each plating process to create the "father," "mothers" and stampers is executed additional mechanical tolerances and microscopic differences will be introduced again resulting in varying levels of intrinsic errors. Each stamper will have its own unique set of errors as a result of the tolerance of punching the stamper and mechanically mounting that stamper into a molding chamber.

Once the molding process begins sources of error are mechanical wear on the stamper (a single stamper can create up to 10^5 impressions), as each disc is stamped the stamper wears, resulting in disc #1 of that stamper being different than disc #10^5 from that stamper. However, if the line is run less than 10^5 discs and the stamper is removed and subsequently placed back into a mold chamber the process of dismounting the stamper, handling, storing, and reinstalling the stamper will introduce mechanical tolerance changes.

Each disc created by the molding process is subject to the feed temperature of the polycarbonate, the heat, humidity and pressure in the mold chamber, the quality of the polycarbonate, and how rapidly the polycarbonate is cooled. The mechanical handling of the separation from the stamper and transfer into the remaining processes can all introduce mechanical stresses and changes that will impact the final error signature of the disc. For example, the speed at which the polycarbonate cools and how rapidly the polycarbonate is pulled from the stamper will create changes in the shapes of the pits and lands, these changes can result in errors. The sputtering processes to apply either the semi reflective material or the fully reflective aluminum also have mechanical tolerances that will impact the thickness of the reflective material as well as the amount of reflectivity across the surface of the disc. Changes in reflectivity of the disc as it is scanned by the laser in the DVD player will impact the error rate of the disc. Bonding the two half discs together introduces potential differences in the run out of the two half discs. Finally the finishing of the label on the top surface of the disc can introduce mechanical stresses that create errors. All of these sources of mechanically induced differences in the finished disc will impact its error rates.

4 Summary

Storing one bit on an optical disc costs $\sim 10^{-13}$ dollars, far less than on most other storage media. In this paper, we propose o-DNA, a cryptographically secure low-cost system for counterfeit deterrence of optical media. We recognize robustness to wear-and-tear as the only design criterion for optical discs that implicates o-DNA's efficiency.

Acknowledgements

We would like to thank Vencil Wells from AudioDev, Dr. Paul Liao from Panasonic Research, Jared Feldner from LeCroy, Dr. Holger Hoffman from Technicolor, Hiroo Umeno and June Dorris from Microsoft XBOX, and Dr. David Heckerman from Microsoft Research for discussions that improved this article.

References

1. G.J. Simmons. Identification of data, devices, documents and individuals. IEEE International Carnahan Conference on Security Technology, pp.197–218, 1991.
2. G. DeJean and D. Kirovski. RF-DNA: Radio-Frequency Certificates of Authenticity. Cryptographic Hardware and Embedded Systems, pp.346–363, 2007.
3. IEEE 1363-2000: Standard Specifications For Public Key Cryptography, 2000.
4. Standardizing Information and Communication Systems. 120 mm DVD - Read-Only disc. Standard ECMA-267. 3rd Edition, 2001.
5. Audiodev CAT300 DVD Reference Player. `http://www.audiodev.com/?id=2088`.
6. O. Slattery, et al. Stability Comparison of Recordable Optical Discs – A Study of Error Rates in Harsh Conditions. Journal of Research of the NIST, Vol.109, no.5, pp.517–524, 2004.