Factors in a Security Printing & Imaging Based Anti-Counterfeiting Ecosystem

Steven J Simske, Margaret Sturgill, Jason Aronoff, Marie Vans; Hewlett-Packard Labs; Fort Collins, CO, USA

Abstract

Security and forensic printing are needed to connect a physical object to the infrastructure—servers, databases, services, etc.—that is necessarily deployed for the "downstream" aspects of an anti-counterfeiting ecosystem. These aspects include investigation (secret shopping, evidence gathering, and analytics) and prosecution. For many branded products, including those of our company, the overwhelming majority of counterfeit goods are produced by a few large-scale counterfeiting operations. Therefore, an effective security and forensic printing campaign will be targeted at discovering the presence of counterfeiting in the supply chain as fast as possible, determining the size of each counterfeiter, and prioritizing the evidentiary and prosecution plans to eliminate the largest counterfeiters as fast as possible. This paper addresses the factors to be considered in successfully defining an effective security and forensic printing campaign, and early approaches to modeling and simulation of an overall ecosystem to optimize the campaign. Broadly, the following topics are of importance: (1) cost function; (2) input parameters; (3) devices available for deployment; and (4) system outputs. We also discuss the manner in which the solution can be deployed for products with widely different supply chain, counterfeiting and distribution requirements.

Introduction

Security printing is printing concerned with embedding readable information in a printed area which can later be imaged and recovered [1]. Intentional information includes data, often serialized, embedded in a barcode or other visible "deterrent". "Unintentional", or accidental, security printing is associated with the item-unique interaction of ink with substrate, constituting "forensic" level printing [2]. In addition, large sets of images can be used for what is termed "batch forensics" due to the increased analysis probabilities associated with the analysis of large image sets [3].

Most security printing approaches are based on ad hoc analysis of the absolute effectiveness of a given approach, or even better the relative effectiveness of different security approaches. Because these approaches are often based on single-factor costs, they do not accurately represent the cost—or the value—of the printed marks in the larger "ecosystem"—meaning combined set of operations, or tasks—in which the deterrent is deployed.

In order to address the real value of a deterrent, then, a cost function taking into account the overall system costs of deployment and use of a security mark must be defined. These cost functions are used to optimize the return on investment (ROI) of the ecosystem. Overall cost of intervention (additional costs for security features, secret shopping, etc.), time to response, time to capture and asset inertia—making best use of tools already available in the supply chain and at the point of sale—are key elements of the cost function.

In such a cost model, input parameters include counterfeiting rate, which can be assessed by a number of indicators: unexpected rebate volume, lower-than-expected sales, etc. The number of large counterfeiters is important, and this can be addressed by existing (image based forensic) means [3]. Layout of distribution network and how product sampling is achieved are other inputs.

Device selection is another crucial part of modeling the ecosystem. Data-gathering devices range from the expensive and specialized—such as RFID readers and USB-powered microscopes—to the inexpensive and commonplace—such as mobile cameras. The trade off between these devices for security, reliability, and security payload density plays a role in how the overall ecosystem recommendation is made.

Output from the model is the deployment recommendation for the brand owner, which can be complicated by the need to accommodate multiple products simultaneously. Typical recommendations focus on how and when to deploy mass serialization, authentication, inspection, forensics and spot checks in the supply chain. Product-specific elaborations include the cost of counterfeiting—lost sales, returns, future lost sales, liability, etc.—in addition to considerations of what percentage of the counterfeiting is actually addressable and/or preventable. Further considerations include the cost of recall and the finality of intervention—counterfeiters who are simply slapped on the wrist are likely to be ambidextrous enough to use the other wrist to make fake products!

Ecosystem Model

The overall ecosystem being modeled is heavily dependent on the imaging (reading) devices deployed. Table 1 overviews some of the devices deployed along with their locations, agents using them, and the cost of using them (fixed and per-use). Five devices are considered: inspection cameras placed on the manufacturing/printing line; barcode readers used at distribution location, supply chain nodes, and/or point of sale; scanners (including all-in-one devices) used throughout the supply chain; mobile cameras used by end users; and forensic imagers such as those described in [2] used by knowledgeable agents throughout the supply chain.

Inspection cameras have high fixed costs but very low per-use costs thereafter. Barcode readers are similar—we used pricing for a 2D barcode reader since these marks are more relevant for current supply chains (in which 2D marks are used for mass serialization) and indeed at point of sale (two of the largest US retail brands—Target and Wal-Mart—are fitting all stores or all new stores, respectively, with 2D barcode readers).

Scanners, including all-in-one devices, multi-functional printers (MFPs) and copiers—are inexpensive to purchase, but require more time—and typically expertise/training—to use for inspection, authentication and other imaging tasks. Mobile cameras, on the other hand, are ubiquitous, and as such require no fixed cost for use—although the per-use cost is non-zero, since there are significant incremental costs over other imaging devices. Mobile camera usage is also tied to incentives to customers to use them; for example, couponing, gaming and other loyalty programs. These incur some costs for the brand owner. Additional costs are incurred in the development of imaging software with broader capabilities to enable the support of the plethora of mobile camera technologies.

Device	Factor	Data
Inspection Camera	*Location*	Manufacturing/printing line; Re-packaging centers (if applicable)
	Agent(s)	Manufacturer, distributor
	Fixed Cost	$4000.00
	Per Use Cost	$0.05
Barcode Reader	*Location*	Distribution centers; supply chain nodes; point of sale
	Agent(s)	Distributor, retailer
	Fixed Cost	$1000.00
	Per Use Cost	$0.10
Scanner	*Location*	Throughout supply chain—especially at the retailer
	Agent(s)	Retailer, inspector, some customers
	Fixed Cost	$100.00
	Per Use Cost	$1.00
Mobile Camera	*Location*	End users / customers
	Agent(s)	Customers
	Fixed Cost	$0.00
	Per Use Cost	$0.05
Forensic Imager	*Location*	Knowledge agents; including auditors and recall managers
	Agent(s)	Inspector; forensic agents
	Fixed Cost	$50.00
	Per Use Cost	$1.00

Table 1. Reading devices, factors in the cost model associated with each (location, agent(s) using them, and fixed and per-use costs), and the values associated with each factor.

As an example of the reading costs, we consider here two scenarios: mass serialization for point-of-sale $5 product (called P) validation and retailer validation of an over-the-shelf $50 value medical (called M) product. We compare equal costs, so in this case we assume there are 10^6 units of P and 10^5 units of M. The total product values are thus $\$5x10^6$. For P, the costs are for the inspection camera ($\$4000 + \$0.05x10^6$) and for the mobile camera imaging ($\$0.05x10^6$), which sum to $104k. For M, the costs are for the inspection camera ($\$4000 + \$0.05x10^5$) and for the scanner ($\$1.00x10^5$), which sum to $109k. Thus, the costs are roughly the same for equivalent values of products M and P. There is one difference, however: full compliance is expected in the case of product M but full compliance is not expected P (even though the costs, in general, cannot be recovered when compliance is less than 100%).

These costs can be broke out further. More generally, the costs in the ecosystem are:

$$\text{Cost} = P_m*W_m*C_m + P_i*W_i*C_i + P_a*W_a*C_a + P_r*W_r*C_r \quad \textbf{(1)}$$

where $\boldsymbol{m}$ represents the costs in the manufacturing/production process, $\boldsymbol{i}$ represents the costs in the imaging process, $\boldsymbol{a}$ represents the costs in the authentication process, and **r** represents the costs in the recall process. P is the probability of using each of these costs and W is a weighting factor to account for differences in how the costs are incurred. For example, for the imaging costs of product P, $W_i = 1/P_i$, since the costs for developing the image analysis systems and deploying the customer incentive programs is incurred regardless of the overall use rate by the customers.

Sensitivity of the Model

Equation 1 provides a general cost model for the deployment of security and forensic printing information. It should be noted that these models are highly sensitive to modest changes in per-use cost, since they are typically deployed for large-volume products. In the case of product M, for example, simplifying the process for scanning so that the per-use cost drops to $0.50/item drops the overall cost for $\$5x10^6$ worth of product to just $59k, making it far more cost-effective than for an equivalent worth of product P.

In general, then, the sensitivity of the model is greatest where the first derivative of the costs/unit are highest—that is, where $\partial C_x/\partial n$ is maximal, subject to $x \in \{\boldsymbol{m},\boldsymbol{i},\boldsymbol{a},\mathbf{r}\}$ and n=number of units. To identify the maximum sensitivity, the overall ecosystem must be carefully considered. If, for example, $W_x \propto 1/P_x$, then relative sensitivity of P with respect to W must be multiplied by $\partial C_x/\partial n$ to obtain the overall sensitivity. This means that $\partial C_i/\partial n$ for product P is 0.0; in other words, it is cost insensitive (unless software system or customer incentive costs can be reduced in the large).

The Model in Action: Recall

In order to bring into play the full cost model described by Equation 1, we consider the costs involved in multiple stages of several workflows, the most important of which is recall (removal of product from the supply chain), since it is the workflow that incorporates all elements of the model.

For compliance and quality assurance (QA), often the cost involved is solely in the manufacturing/production line, and so the overall cost model reduces to:

$$\text{Cost}_{\text{compliance,QA}} = P_m*W_m*C_m \quad \textbf{(2)}$$

For supply chain analytics, imaging costs will be incurred by, minimally, some distributors and retailers:

$$\text{Cost}_{\text{supply chain analytics}} = P_m*W_m*C_m + P_i*W_i*C_i \quad \textbf{(3)}$$

The cost of determining the presence and level of counterfeiting involves an additional authentication cost:

$$\text{Cost}_{\text{authentication}} = P_m*W_m*C_m + P_i*W_i*C_i + P_a*W_a*C_a \quad \textbf{(4)}$$

Finally, if product recall needs to occur—due to counterfeiting, tampering, product repackaging, etc.—an additional recall cost is incurred. Thus, recall in general is governed by Equation 1 and the workflow in Figure 1.

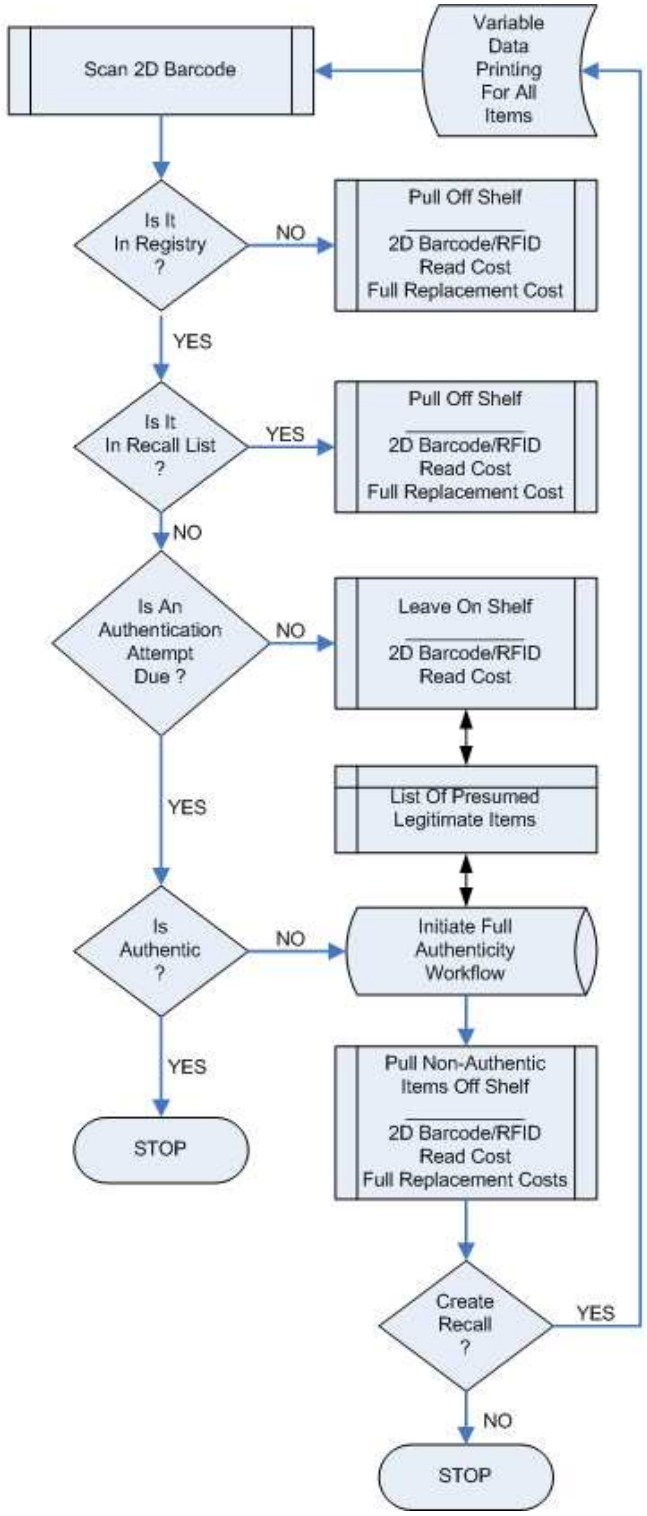

Figure 1*.Block diagram of the recall ecosystem.*

In Figure 1, the security mark deployed is the familiar 2D barcode, of which the Aztec [4] and DataMatrix [5] are familiar examples. At any of various points in the supply chain, these 2D barcodes are read—by inspection cameras in the manufacturing line; by barcode readers, scanners and/or mobile cameras during the routine imaging; by barcode readers, scanners, mobile cameras and/or forensic imagers as part of the authentication; and by any or all of the imaging devices (Table 1) during the recall, depending on the nature of the supply chain threat.

Generally, barcode readers will be sufficient for imaging. However, during recall, occasional authentication (reading of unique imaging information) will need to be performed even where all barcode reads are apparently authentic. In this cases, additional security marks [1] or forensic analysis of the printing itself [2][3] will be needed. This will add to the per-use costs, but the relatively low P_a and P_r values will keep the overall ecosystem costs from rising. We now illustrate this by example, referring to Figure 1.

When the need for product recall is defined, the package is scanned with the appropriate barcode reader at any node in the supply chain. Every unit must be imaged. Each barcode read belongs to one of these three classes:

1. Legitimate barcode number, not repeated elsewhere
2. Legitimate barcode number, repeated elsewhere
3. Non-legitimate barcode number

Those belonging to class (1.) are the only ones that can be safely left on the shelf under any conditions, but further authentication must take place to achieve statistical confidence in them. All of class (2.) must be removed, even if they belong to an otherwise authentic batch (implying they had simply served as the source of one or more legitimate numbers for other units belonging to this class), simply because they are suspect. All of class (3.) must be removed, as they are certainly counterfeits.

In order to leave a batch of class (1.) barcodes on the shelf during a recall, however, we must sample N samples out of batch size M, with probability of a false positive P_{FP} for the authentication known from previous analysis, such that:

$$(M/N) * (P_{FP})^N < P_{FS} \quad \textbf{(5)}$$

Where P_{FS} is the forensic security probability, or the required maximum probability of any samples in the batch being counterfeit. For example, if $P_{FP} = 0.001$, $P_{FS} = 10^{-12}$, and $M = 10^6$, then solving for N, we see that only 7 samples must be checked to have confidence that less than 1 in 1/ P_{FS} of these samples are counterfeit—in spite of a relatively modest value for P_{FP}. This is because $(M/N) * (P_{FP})^N = 1.4 \times 10^{-14}$, which is less than P_{FS}.

Thus, a relatively modest cost is incurred for forensic analysis of the entire batch; that is, $P_a*W_a*C_a$ is much less than the first two costs, $P_m*W_m* C_m + P_i*W_i*C_i$. The authentication costs are indicated in the lower part of Figure 1, where the decision box "Is An Authentication Attempt Due?" is answered by the sampling frequency N/M determined from Equation 5. The full authenticity workflow is thereafter governed by P_a = (N/M), which in the above example is a modest 7×10^{-6}.

However, the right column in Figure 1 describes a set of costs not yet discussed. These are the recall costs, or $P_r*W_r*C_r$. The

per-use costs for recall are at first glance high: every non-authentic item must be pulled off the shelves. But, we have shown above how to contain these costs by quickly assigning batches to one of three classes. If any items in a batch are assigned to either class (2.) or class (3.), then the entire batch is disposed of with the concomitant economy of scale. Note that a "batch" can be a carton, pallet, shipping container, or other logical unit, based on the relative costs of the items and the authentication.

Conclusions

This paper introduces a simple, but highly adaptable—model for determining the costs involved in a printing-oriented security and forensic ecosystem. Most of the concepts overviewed are equally applicable to non-printing based ecosystems; for example, RFID and other sensor-based ecosystems. By allowing a term for the percentage of samples analyzed during manufacturing, imaging, authentication and recall, along with a weighting term to incorporate the realities of overall versus per-use costs, the model is not limited to linear combinations of costs.

The paper also provides a breakdown of the reading costs involved for five different types of imaging devices, at each of the four workflow stages described. These costs are readily incorporated into the overall cost model.

We also provide a description of how to perform sensitivity analysis on the model. In order to minimize system costs, both the cost sensitivity and the relationship between the probability and weighting factors must be considered.

We then showed how barcode reading by itself can be used to quickly assess and assign products to three actionable classes in the case of recall. We showed how recall costs can be contained to reasonable levels through consideration of the full model.

Future Work

The model presented is not complete. Future work will focus on tying the reading costs, as outlined in Table 1, into the cost structure defined in Equation 1. In so doing, a more substantial, and closed-formed, solution to the sensitivity analysis may be discovered.

The reading costs, fixed and per-use, also need to be broken down for different device configurations. The examples provided herein indicate that fixed costs, such as those for expensive inspection cameras, are diminishingly small compared to the accumulated per-use costs when large numbers of products are involved. The results also indicate that it is, logically, safe to increase the per-use costs in proportion to the relative cost of each item. Future work should further elaborate on this, and also consider the differences between revenue and margin for the different products.

Finally, Equation 5 emphasizes the need for quick assessment of multiple-unit validation with "random" full authentication of units with frequency (N/M). Figure 2 shows an example of how variable data printing can be used to enable such a quick validation/authentication of multiple units. Figure 2 represents 6 small packages lined up in a carton or box. Each package has a static set of printed colors in one part of the set of color lines printed on the side of the box, allowing rapid validation. Each package also contains variable data with low P_{FP}, allowing for quick authentication of the lot through sampling a small N. Future work will focus on additional ways variable data printing can be used to reduce the overall ecosystem costs.

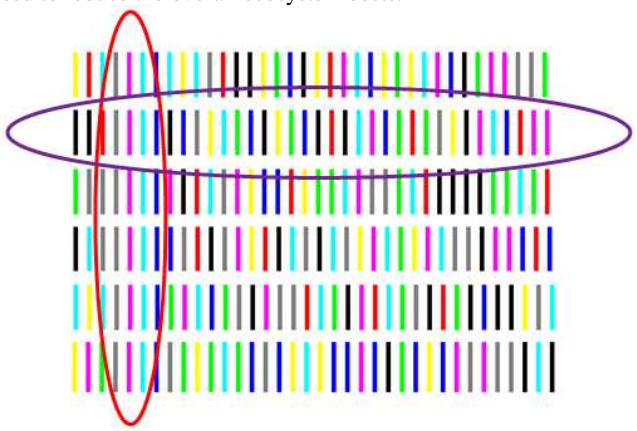

Figure 2*.Example of the use of static colors on the side of packaging for quick "validation" of multiple units (encircled by vertical/red oval) and dynamic colors for individual item authentication in accordance with Equation 5 (encircled by horizontal/purple oval).*

Acknowledgements

The authors gratefully acknowledge Guy Adams, Paul Everest, George Guillory and many others involved in the security and forensic ecosystems of which this paper merely scratches the surface.

References

[1] S. Simske, M. Sturgill, G. Adams and P. Everest, "Document imaging security and forensics ecosystem considerations," Proc. ACM DocEng 2010, in press (2010).

[2] S. J. Simske and G. Adams, "High-resolution glyph-inspection based security system", Proc. IEEE ICASSP, pp. 1794-1797, 2010.

[3] S. Simske, M. Sturgill, P. Everest, and G. Guillory, "A system for forensic analysis of large image sets," Proc. IEEE WIFS 2009, pp. 16-20, 2009.

[4] ISO/IEC 24778:2008, "Information Technology -- Automatic identification and data capture techniques -- Aztec Code bar code symbology specification, International Organization for Standardization, Geneva, Switzerland, http://www.iso.org/iso.

[5] International Standard ISO/IEC 16022, "Information Technology—Automatic Identification and Data Capture Techniques—Data Matrix Bar Code Symbology Specification," International Organization for Standardization, Geneva, Switzerland, http://www.iso.org/iso.

Author Biography

Steven Simske is a Distinguished Technologist in the Print Production Automation Lab (PPAL) in Hewlett-Packard Labs, and is the Director and Chief Technologist for the HP Labs Security Printing and Imaging program. Security printing and imaging provides brand differentiation, brand protection and anti-counterfeiting through novel security algorithms, printing approaches and image analysis technologies. Steven has 35 US patents and nearly 250 peer-reviewed conference and journal publications..