Tracing the Source of a Shredded Document

Jack Brassil

HP Laboratories, Palo Alto CA 94304, USA
email:jtb@hpl.hp.com

Abstract. Consider two ordinary, seemingly identical plain paper shredders labeled A and B. When each of the two shredders is fed a blank sheet of paper, the resulting remnants are sufficiently similar that they are indistinguishable upon visual inspection. Now suppose that one of the shredders has been modified to introduce imperceptible variations in the size of each remnant it cuts. One of the shredders is then selected at random to destroy a blank sheet. By examination of the resulting paper remnants, can one determine if the sheet was shredded by A or B?

In this paper we show how information hidden in the size and shape of shredded page remnants can be used to reveal the identity of the device used for shredding. We describe means for modifying shredders to introduce this hidden information. Experimental results reveal that properly embedded information can survive the severe nonlinear distortions introduced by the mechanics of paper shredding. Finally, we consider the question of whether paper shreds could reveal shredder identity even in the absence of device modifications.

Keywords: jigsaw puzzle reconstruction, machine identification, information hiding, detection theory, image analysis, forensic science

1 Introduction

As long as there is need for either classified, limited distribution or secure paper documents [1], so too will there be need to destroy them beyond the point of practical reconstruction. The attention paid to destruction and disposal practices varies greatly with the need for document security, and the willingness to pay for it. The United States Department of Defense publishes requirements for secure plain paper shredders, and these requirements can be realized at relatively low cost. Secure document disposal is more often a matter of time and vigilance than money. As a result the humble plain paper shredder has become a fixture in environments where the security requirements for waste documents are relatively modest.

It is often the case that the actual distribution of a limited distribution document is of as much interest as the information contained in the document itself. While it might not be possible to know what a recovered, shredded document said or who read it, it is of some interest to know by whom, or where, a document was ultimately destroyed. This is particularly the case if there is suspicion that a sensitive document managed to find its way to an unintended recipient.

It has long been recognized that physical devices routinely leave telltale identification marks or 'fingerprints' of their use [2]. These fingerprints can be unique even if unintended, as is the case if introduced solely by the natural variations in the electrical or mechanical tolerances of device components. As an example, two identically manufactured plain paper copiers will routinely be found to have slightly different magnifications. This subtle observation might enable an examiner armed with an original document to help identify the particular device responsible for creating a plain paper copy. For that matter, at a sufficiently microscopic level no two sheets of paper appear identical, and this observation can be used to distinguish between two seemingly identical documents.

Reconstruction of fragmented documents has long been of interest to researchers in both computer science and the social sciences. Ancient manuscripts such as the Dead Sea Scrolls have been painstakingly reconstructed [3]. In the case of the scrolls, who authored and possessed the documents is as central a question as the reconstructed text itself. Since the earliest days of digital computers, computer scientists have considered the problem of the computer-assisted reconstruction of both 2-dimensional and 3-dimensional jigsaw puzzles. Reconstructing a document from fragments may be considered an unusual but highly specialized case of puzzle reconstruction [4], [5].

The generalized reconstruction problem remains formidable, and we do not consider that problem here. It is worth noting that the attacks on this problem that we are aware of make use of the fragment content in addition to its shape. Instead we consider the simpler problem of whether the remnants of a shredded document can be associated with the shredding device it passed through. We have no knowledge that the identification problem examined in this manuscript has been considered by other researchers. Yet a great deal of effort has been directed to the recovery of supposedly 'deleted' documents from computer disk drives, as well as tracking ownership and distribution via digital watermarking. However the recent investigation into fraudulent accounting practices at a major US corporation, and the subsequent assertion that paper documents were wrongfully destroyed by its accountant, has encouraged us to consider this question.

The remainder of this paper is organized as follows. Section 2 briefly reviews the mechanical operation of common office shredders. Section 3 describes how to hide identifying information in a shredder, and the experiments we performed to gauge our ability to detect that hidden information. The next section discusses alternative approaches to embedding identifying marks in a device. Section 5 develops a mathematical model for detecting marks in the presence of the severe 'noise' introduced by the shredding process. Section 6 analyzes the probability of distinguishing between two shredders without modification to either shredder, and the final section summarizes our results.

2 Basic Operation of a Plain Paper Shredder

Though secure techniques for destroying paper are easily realized (e.g., burning), it is a simple fact that those techniques are used only in the most secure

environments. Most of the rest of us settle for the low cost and convenience of common office-grade electronic shredders, a market exceeding US $100 million per year. Practical document destruction appears premised on the likelihood that few will bother to go to the effort of reconstruction, and also that other enterprise security risks (e.g., network attacks) are a larger threat.

Though the mechanics of document shredding vary dramatically, shredders tend to fall into a small number of distinct security classes. Minimally secure shredders for office applications known as *strip shredders* cut paper only along its length (i.e., perpendicular to the shredder mouth). The resulting shreds form long narrow strips, with common widths ranging from 1/8 to 5/16 inch. Because of the requirement to destroy multipage documents bound by paper clips or staples, the typical cutting assembly is surprisingly strong. A common cutting mechanism employs two arrays of rapidly rotating, ribbed metal bands (caterpillar tracks or treads) whose width defines the strip width. These opposing band arrays come together at an acute angle to form a V-shaped mouth with adjacent bands interleaved. The effect of this tearing action will be observed by examining strip edges in figures later in this manuscript.

The *crosscut shredder* cuts both lengthwise and widthwise to achieve additional security protection. Crosscuts occur relatively infrequently, often being made only every few inches of page length. A relatively recent innovation is the *confetti* shredders, which crumples crosscut remnants. One of the oddities encountered in the marketing of shredders is the relative prominence of convenience over security; crosscut shredders advertise that densely packed remnants offer the convenience of fewer disposal trips, stripcut shredders advertise fast cutting, and confetti shredder ads go as far as suggest that the remnants are to be used for packing materials.

To establish a security baseline, the US Department of Defense has specified that Class I secure shredders crosscut documents to yield remnant sizes no larger than 1/32 inch wide by 1/2 inch long [10]. Note that a 10 point character (approx. 10/72 inch) would be vertically cut several times by such a device, making this presumably adequate to deter attempts at reconstruction. Such a standard, of course, does not ensure that secure papers are destroyed properly. Documents are highly portable, and there are many environments where careful document security is impractical, such as in battlefield settings or mobile workplaces, including airports and hotels.

3 Experimental Setup and Results

Prior to plunging into a discussion of the experiments we performed, let's begin by addressing the most basic question: "Is shredder identification feasible?" The answer is clearly *yes*, as illustrated by the following pedagogical construction. Suppose strip shredder A cuts each inch of page width into four identical pieces, each 1/4" in width. Next suppose that shredder B cuts each inch of page width into four pieces, two 1/8" strips and two 3/8" strips. Then given the remnants of a shredded page, at a glance one can obviously determine whether the result-

ing page was destroyed by A or B. This leaves the remaining question: "Can shredder identification still be performed if the cutting width difference of strips is sufficiently small to be undetectable by human observation of the page remnants?" The remainder of this paper addresses just this question, and identifies just how small such a difference can be. We will show that, given a sufficient number of samples to test, the difference in widths that can be detected can be of the same order as the 'noise' introduced in remnant size due to the tearing action of the shredding process. Even more remarkably, we will go on to consider whether the natural variations that occur in the manufacture of cutting assembly components is enough to permit an observer to identify an otherwise unmodified shredding machine.

Anecdotal evidence suggests that shredders are among the most misused yet least maintained of all electronic office equipment. Rather than consider a new piece of office equipment, we sought out a stripcut shredder that had received considerable use over a period of years, while receiving no maintenance whatsoever. Our objective was to ensure that our results would be robust to the degenerative effects of machine aging and general misuse. We selected a Panasonic Compact Shredder Model MPS20 with a 3 mm (0.118 in.) strip width, a device meeting ordinary office security requirements. We tested the device with blank sheets of 20 lb. ($75g/m^2$) multi-purpose office stock paper. Colored paper was chosen to obtain adequate contrast during scanning, and avoid the occurrence of optical edge effects produced when scanning white paper. Each sheet was separately fed into the shredder mouth, though it is likely that most samples recovered in practice would not have been destroyed one page at a time.

After selecting a shred whose width was to be measured, grayscale (8-bit) scanning was performed at 600 dpi on an HP Officejet R80xi flatbed scanner. Figure 1 shows a typical resulting image of a strip cut.

Fig. 1. A strip cut.

Observe how the jagged nature of the strip edge shows how the page is torn rather than cut. In some samples, particularly careful observation can reveal indications that the remnant passed through a cutting assembly using ribbed bands, as the resulting strip outline take on a slight crosshatched appearance, as in a sketch of a railroad track or a corral fence.

Scanning requires confronting a number of minor practical considerations. If one was to perform a smoothing operation on the jagged strip edge, one would

find find that opposite edges of the strip to be very nearly parallel. But the nature of a thin flat strip is to tend to curl (e.g., like a human hair). This curling action tends to cause a strip to not lie flat when placed on a flatbed scanner without being secured. Hence, the paper edges in scans of individual strips appear 'wavy'. Figure 2 shows this waviness of the strip when a grid is superimposed over the of the image in Figure 1. Ultimately we found it convenient to secure strips to a rigid background prior to scanning to avoid an excessive amount of this waviness due to curling.

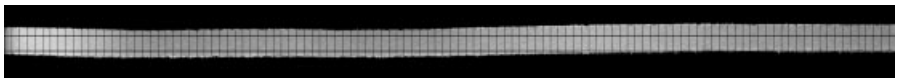

Fig. 2. A strip cut with overlayed grid.

Scanning a strip generally produces a skewed image of a strip, that is, one having a rotation with respect to the horizontal axis. Note how the strip in Figure 1 is not sitting horizontally across the page, but is slightly tilted. But even if modest skew occurred ($\theta = \pm 2.5°$) then the error in width calculation would only be .118 in $\times$ 600 dots/in. $\times \cos 5° = 0.3$ dots, or less than 1 pixel in error. Relatively speaking this is a minor source of error, so we choose not to bother attempting to correct any skew.

A final consideration in scanning paper strips was to ensure that the strip was oriented such that the movement of the scanner's light bar traveled along the strip length, rather than the strip width (to be measured). Otherwise, shadowing and other optical edge effects in the scan created problems in detecting the strip edge, making width measurement unnecessarily difficult.

Following scanning, each image was processed using a collection of software tools to prepare for a width measurement. We used publicly available tools found in [6], occasionally supplemented by a comparable feature available in the shareware tool *xv* version 3.10a. Image preparation typically involved:

1. *cropping* to reduce image size,
2. a 90% *rotation* to approximately horizontal,
3. *binarization* (i.e., color or grayscale image to binary),
4. *despeckling* to remove noise introduced by scanning,
5. *inversion* to a white strip on a black background, and
6. *edge detection* (to extract the noisy cut edge).

Figure 3 shows a typical result of this process. Often, a final cropping was helpful to remove artifacts such as the detected left edge in the figure (corresponding to the strip top or bottom and of no immediate interest).

The combination of skew and waviness caused us to abandon our initial simple approach to measuring the strip width, which was roughly to be as follows.

Fig. 3. Noisy cut edges whose separation is to be measured.

Suppose that we binarized and inverted Figure 1, such that we created a binary image, where the strip was white and the background black. We would then create a horizontal projection profile, that is a histogram of the number of "on" bits per horizontal row. Ideally, a projection onto the vertical access would show a distinct transition at the edge of each strip, enabling us to measure the strip width by finding the difference between the transitions (in pixels). But our experiments showed that waviness and skew easily caused that distinct transition to blur over approximately 5 or 6 pixels, producing too much variability in width measurement.

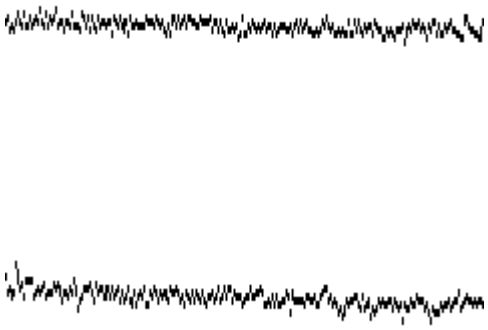

Fig. 4. A closer look at the detected edges of Figure 3.

To circumvent these problems we found it necessary to write a dedicated program to precisely estimate the strip width. This program measured the exact

distances between the edges detected in the (rotated, binary, despeckled, inverted) image for a large number of points along the strip's length. A histogram of the measured strip width is shown in Figure 5. The distance was estimated at 5601 points that our algorithm deemed sufficiently free of noise to obtain an accurate measurement. This represented about 90% of the points across the entire strip length. The width of this strip was found to have mean 65.85 pixels, and variance equal to 9.045 $pixels^2$ ($\sigma \approx 3$ pixels, or 1/200 inch). We will develop a detailed analytical detection model shortly, but for now our intuition suggests that we would be able to reliably detect the difference between two strips differing in width by $2\sigma \approx 1/100$". And as we had hoped, that difference is small enough to be effectively imperceptible to all but the most astute human observer.

4 Alternate Techniques for Embedding Identifying Information

So far we have limited our discussion to a single means of hiding identifying information in a single type of stripcut shredder, namely by embedding information in the size of remnants. There are several ways in which this technique can be realized, among the most straightforward being using cutting bands with imperceptibly different widths. In the case of the device we used, the manufacturer was already known to have cutting components producing 3 mm. and 4 mm. cuts. Note, however, that the experiments described here are suggesting that a much smaller width difference, such as 0.25 mm, would most effectively balance detection accuracy and imperceptibility.

Other means of embedded information are possible, and in some cases may be preferred. For example, other features such as the overall shape of a remnant may be useful for hiding identifying marks. Consider the effect of carving small notches or indentations in cutting bands, producing small but detectable imperfections in the cut edges of remnants occurring at well known intervals. Such a technique would in principle be effective on a variety of shredder types, including crosscut shredders. Though we have not performed experiments demonstrating that this is possible, the experiments we have done suggest that we may be capable of detecting notches of perhaps .005 inch in depth. A sequence of such notches could be used to encode a unique identifier using any number of techniques, such as their relative positions (analogous to pulse-position modulation in digital communication systems). Though it is possible to imagine modulating information in the path of a cut, an inexpensive mechanical design might be difficult to realize. But if successful, such an approach may be capable of embedding far more information in a shredded page than would be possible by modulating strip widths.

It is worth noting that the analysis of shred sizes is only one technique that might be useful in a machine identification. In practice, there are many artifacts of the shredding process that can help the identification of a shredding source. As an example, paper dust within the shredder enclosure can be examined for

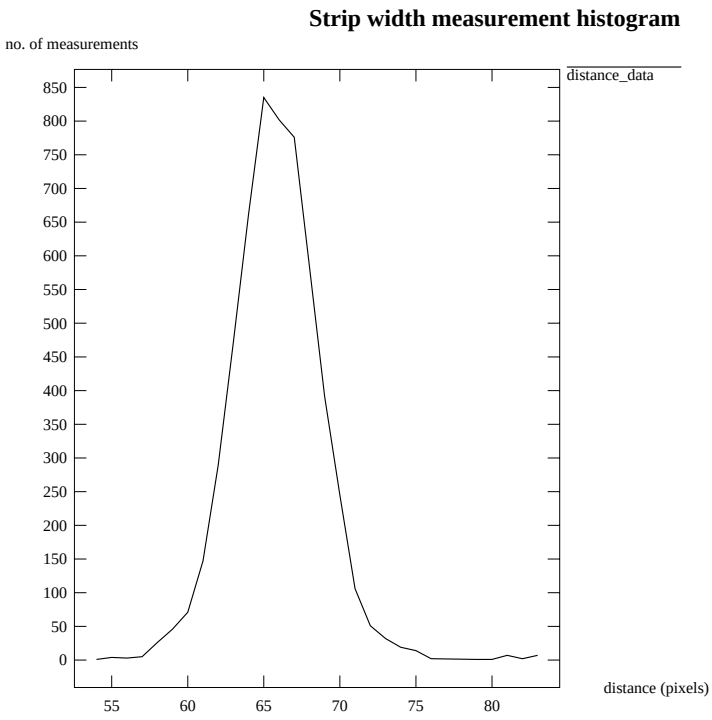

Fig. 5. Width measurement histogram.

fibers or dyes that are consistent with those found in shred waste. Even more simply, paper length might be revealing when compared to the operating profile of the device environment. For example, a legal department might produce more 14" paper than a human resources department.

We have not had an opportunity to examine the effects of device aging on detection accuracy. Shredders are notoriously some of the most abused and poorly maintained of office equipment, so it is likely that equipment will show the affects of aging. Dulled cutters are likely to increase tearing and produce jagged cuts, likely making device identification harder or even impossible. On the other hand, other artifacts that could be helpful in the detection process might in fact be amplified by aging or misuse. For example, one could imagine distinguishing between a new and an old shredder by simply examining the jaggedness of paper tearing at a strip edge, which is likely to increase with device use. A device's signature may in fact become more pronounced with age.

5 Analytical Detection Model

We next develop an analytical detection model to evaluate our ability to trace shredder identity. Let's begin with the simple case of deciding whether a single strip X was produced by shredder A or B. Label the sample strip width x, and assume that the corresponding strip widths produced by machines A and B are known to be a and b, respectively. We assume that each shredder has been modified to make imperceptible increases or decreases to each strip's width, such that $a \neq b$.

Let H_a (H_b) be the hypothesis that the sample passed through machine A (B). Then, if it is equally likely that the sample was destroyed by either machine, the decision rule we would implement is as follows:

$$if \quad \begin{array}{l} |x-a| < |x-b| \ : \ decide \ \ H_a \\ |x-a| > |x-b| \ : \ decide \ \ H_b \end{array} \tag{1}$$

Now let's move on to the general case of identifying a device's unique k bit identifier, permitting us to distinguish between each of 2^k shredders. Suppose that our sample collection $\overline{X}$ contains the (ordered) shredded output of an entire page, and $X_i : i = 0, 1, 2, \ldots, N+1$ is the i^{th} strip from the left that exited the shredder. Let $x_i, i = 1, 2, \ldots, N$ be the width of the i^{th} strip, as measured by the process discussed in Section 3. Note that we are excluding the strip at each page edge, which we assume to not have two cut edges.

We will hide a unique identifier in the page remnants produced by each machine by embedding a binary code in the remnant strip widths. To make detection as simple as possible, 1 bit will be encoded by using two adjacent strips, in the following fashion:

- A "zero" is embedded by increasing the width of odd-numbered strip i by distance Δ, and decreasing the width of strip $i+1$ by Δ.
- A "one" is embedded by decreasing the width of odd-numbered strip i by distance Δ, and increasing the width of strip $i+1$ by Δ.

We will refer to the hypothesis corresponding to each of the above modifications as H_0^i and H_1^i, respectively. Note that the sum of the widths of any two adjacent strips remains the same after this modification, as does the sum of all strip widths.

Since a single bit of identifying information is embedded in every other strip, an 8-1/2 in. paper sheet passing through a shredder making a typical $3/16 = 0.1875$ in. wide cut would permit embedding $\lfloor \frac{8.5/0.1875-2}{2} \rfloor = 21$ bits in each page, where we subtract 2 bits corresponding to the possibly incomplete strips at the paper edges.

Note that the decision rule analogous to (1) would require information about the widths of each strip for each of as many as 2^k shredders. This is the case because the strip widths of each machine may be different prior to modification, as a result of natural variations in manufacturing. However this decision rule can be greatly simplified if we make the following assumption. If the widths of

both the 'narrow' and 'wide' strips cut by all modified machines are identical, then we can simply compare the measured widths of two adjacent strips with the following decision rule:

$$if \quad \begin{array}{lll} x_i < x_{i+1} & : \; decide & H_1^i \\ x_i > x_{i+1} & : \; decide & H_0^i \end{array} \qquad for\ i\ odd \tag{2}$$

In practice, of course, manufactured components are not identical (a fact we will exploit shortly) so we can not expect the strip widths to be such. However if the magnitude of the variation in manufactured components is small relative to the displacement Δ, this assumption is reasonable. The practical benefits of making this assumption are large, since (2) frees us from the requirement of access to all machines to perform measurements.

We would next like to calculate the probability of making an error in detecting a single encoded bit. To do so, we observe that there are many sources of 'noise' affecting our detection accuracy, including imprecision in component manufacture, variability due to paper types, machine aging, and measurement errors. Suppose that we model the composite affects of all these error sources as additive noise through a collection of $n/2$ (n even) independent and identically distributed (i.i.d) random variables $n_i : i = 1, 2, \ldots, n/2$, each with zero mean and variance σ^2. That is,

$$\begin{array}{l} H_1^i : x_i = a_i + n_i \\ H_0^i : x_i = b_i + n_i \end{array} \qquad for\ i\ odd \tag{3}$$

Assuming that each shredder was equally likely, we can write the detection error probability as

$$\begin{array}{l} P_{error} = 1/2P[x_{i+1} - x_i > \Delta | H_1^i] + 1/2P[x_{i+1} - x_i > \Delta | H_0^i] \\ \qquad = P[x_{i+1} - x_i > \Delta | H_1^i] \end{array} \tag{4}$$

If we further assume that the noise variables are Gaussian, then the conditional probability density function $x_{i+1} - x_i | H_1^i$ is a normal random variable with mean $-\Delta$, and $x_{i+1} - x_i | H_0^i$ is a normal with mean Δ. We can then write the probability of making a single decision error in (4) as

$$\begin{array}{l} P_{error} = P[n_i > \Delta] = \int_{\Delta}^{\infty} \left(2\pi\sigma^2\right)^{-1/2} e^{\frac{-x^2}{2\sigma^2}} dx \\ \qquad = erfc(\frac{\Delta}{\sigma}) \end{array} \tag{5}$$

To provide a sense of how small a strip width difference Δ is required, if we set Δ equal to the standard deviation of the composite 'noise', than the probability of correctly guessing an embedded bit is $erfc(1) \approx 0.84$.

Given our ability to embed a single bit in every other strip (totaling about 21 bits for an 8-1/2 wide page), how best should we embed a unique identifier? Here we would naturally rely on standard error correction techniques, likely combining binary block coding and interleaving. A properly selected code could easily correct as many as 3 errors in these 21 bits. As a simple illustration, if we interleaved 3 Hamming (7,4) codes we would be capable of distinguishing

$2^{12} = 4096$ uniquely encoded shredders. Using this (naive) scheme, if we set the strip width difference $\Delta = 2\sigma$ as suggested in Section 3, the probability of correctly identifying a shredder among its 4095 colleagues would be

$$1 - P_{error} = [\binom{7}{0}.995^7 + \binom{7}{1}.995^6 \times .005^1]^3 = .998 \quad (6)$$

Let us pause momentarily to revisit some of the assumptions made in arriving to the result in (6). Recall that we assumed that each strip could be associated with the cutting band which created it. That is, not only is each strip indexed, but we can explicitly associate it with a particular cutting band. Note that the shredder operator ordinarily dictates where the paper is fed into the device's mouth. So in practice, achieving this alignment might take some care, and may require embedding marks solely for alignment purposes. Indeed, overhead lost to 'alignment' or 'synchronization' is quite common in detection systems. For example, in principle this alignment is no different than the use of standard center band and edge guard band sequences in 1-dimensional bar code systems such as the Universal Product Code [8].

To this point we have also assumed that we have recovered an entire sample page to test (or more precisely a collection of remnants forming a complete horizontal strip across a page). What if this is not the case? Suppose, for example, we are confronted with a bag of shreds produced by one of two hypothetical shredders A or B, created entirely from blank sheets of paper of unknown original page widths. It appears that we may again be able to identify the shredder used, assuming that the device embedded a fingerprint in a fashion designed to handle this form of remnant recovery. A naive approach to this would be as follows. Machine A embeds no fingerprint (i.e., no modification of the strip widths it produces) while machine B uses the width modulation technique discussed above. Then, upon measurement of large numbers of shreds, the appearance of a bimodal distribution in the width histogram (as in Figure 5) would clearly reveal that the shreds were produced by B.

6 Detecting a Native Fingerprint

Is it possible to determine whether one of two shredders destroyed a document when neither shredder has been modified to embed a fingerprint? We will next show that there is reason for cautious optimism regarding this problem. Once again we limit our attention to a binary decision problem, assume that our recovered sample comprises an entire page, and assume that we have access to the shredders for measurement purposes.

Our ability to distinguish between two shredders rests on our expectation that the cutting assembly components will be manufactured to within some allowable tolerance, and that such imprecision will be of sufficient magnitude so as to be detectable. Suppose that we assume that, due to manufacturing tolerances, the variation in widths of each strip cut by machines A and B can be represented by i.i.d. random variables. Suppose we assume that these variables are uniformly

distributed, and that mechanical tolerances produce a ±5% variation about a strip width mean of 3 mm. (0.118 in.). Then we may write the probability density function

$$f(x) = f_A(x) = f_B(x) = \begin{cases} 1 & .118 - .0118/2 \leq x \leq .118 + .0118/2 \\ 0 & otherwise \end{cases} \tag{7}$$

To distinguish between the 2 shredders, it is necessary to find a detectable difference in the width of the corresponding strip produced by A and B. Let's suppose that we can detect a difference of Δ inch. Do any corresponding strips have at least this difference is size? The density of the random variable $Z = |X - Y|$ is easily calculated for i.i.d. uniform random variables (see [9], p. 190-191) and equals

$$f_Z(x) = \begin{cases} 20 - 200x & 0 \leq x \leq 0.1 \\ 0 & otherwise \end{cases} \tag{8}$$

The probability that any one pair of associated strips differs in width by an amount larger than our detection threshold is

$$P[Z > \Delta] = \int_{\Delta}^{.0118} (20 - 200x)\, dx \tag{9}$$

If we set the detection threshold to $\Delta = 0.005$ in. (half that recommended earlier in this manuscript) then (9) reveals that a given associated pair of strips produced by A and B will have a detectable difference in width with probability 0.1245, or about 1/8. Recall that a full sheet of paper produces about 40 strips. Hence, we would likely encounter about about 5 instances of strip width differences of sufficient magnitude to be detectable. Hence, there is some reason to believe that a correct binary decision could be made under this set of assumptions. However, far more experimentation with actual modified shredders will be required to empirically verify this conjecture, and the assumption of uniformly distributed manufactured metal components is convenient but unlikely. In closing, it is interesting to note that the tolerance we assumed in our calculation was a modest ±5% variation about a strip's mean width. However [10] specifies that the strip width tolerance for Class I shredders is 1/64 inch, or a surprisingly large ±25% variation about the mean.

7 Conclusions

We have provided initial evidence indicating that it is possible to modify document shredders in such a way that analysis of imperceptible variations in paper shreds can reveal the shredding device's identity. Though we omitted the discussion from this paper, there are numerous simple yet effective countermeasures to our approach to machine identification. No doubt the thoughtful reader has already conceived of several, and all are invited to read the best document destruction practices identified in [10] for some additional clues.

We have barely scratched the surface of an investigation into the detection of hardware device signatures. It may well be that the signatures of devices other than shredders, such as copiers and facsimile, will be just as intriguing to identify yet produce more immediate applications. For those interested in pursuing further work on document destruction, it appears that two large questions are worthy of attention. The first is whether it is possibly to establish the 'channel capacity' of information hidden in shredded page remnants. The second, and likely more important question, is to what extent modifications to remnants provide a means of reducing the complexity of the document reconstruction problem.

Acknowledgements - Thanks to the authors of [6] for developing and making public a set of image processing tools useful in our experiments. Thanks also to Adrian Perrig for informing the author of the existence of [11]. The author is eternally indebted to the recently deceased author of [9] for his teaching and guidance.

References

1. van Renesse, R. L.: Optical Document Security. Artech House, Boston (1993)
2. Wagner, N. R.: Fingerprinting. Proceedings of the 1983 IEEE Symposium on Security and Privacy. IEEE Computer Society (1983) 18–22
3. Golb, N.: Who Wrote the Dead Sea Scrolls? The Search for the Secret of Qumran. Scribner, New York (1995)
4. Ogden, J. A.: The Siting of Papyrus Fragments: An Experimental Application of Digital Computers. *Ph.D. Thesis*, University of Glascow (1969)
5. Levison, M.: The Siting of Fragments of Digital Computers. A.D. Booth (Ed.), Machine Translation, North-Holland, Amsterdam (1967)
6. Seul, M., O'Gorman, L., Sammon, M.: Practical Algorithms for Image Analysis: Description, Examples and Code. Cambridge University Press, Cambridge UK (2000)
7. Helstrom, C. W.: Probability and Stochastic Processes for Engineers. MacMillan Publishing Co., New York (1984)
8. UPC Symbol Specification Manual. Uniform Code Council, Inc., Dayton Ohio (1986). See http://www.uc-council.org/reflib/01302/d36-t.htm
9. Papoulis, A.: Probability, Random Variables and Stochastic Processes. McGraw-Hill Book Co., New York (1965)
10. Information Security Team: Terminator VIII: How to Destroy your Classified Materials. Department of Defense Security Institute (1992). See http://www.dss.mil/training/term4/doc
11. Cohen, F.: The Mathematics of Shredding. Proceedings of the 2002 IEEE Symposium on Security and Privacy. IEEE Computer Society (2002)