

Printer and Scanner Forensics
Examining the security mechanisms for a unique interface

by Nitin Khanna et al.

1 de Outubro de 2010

Revisor: Matheus Nagliati — RA 063091
Rodolfo Udo Labsch — RA 086015

1 Visão global

Os autores descrevem o funcionamento básico de um scanner, uma impressora a laser e uma impressora jato de tinta. A partir disso os autores apresentam métodos para identificar a impressora usada, de modo que possamos identificar livros piratas ou documentos falsos entre outros fatores para ajudar na aplicação da lei.

2 Resumo

Existem duas abordagens para identificação: a primeira que trata de características intrínsecas do dispositivo — características únicas de cada modelo, por exemplo pequenos defeitos de fabricação invisíveis ao olho humano — e a segunda para detecção de características extrínsecas — que envolve imprimir uma marca d'água única na impressão. Para a aplicação de ambos os métodos temos que entender pelo menos o funcionamento básico de um scanner e de uma impressora.

Um scanner é composto de uma luz refletida no documento, que é capturada por sensores luminosos entre cada movimento do motor de passo. Esses sensores de luz apresentam vários defeitos que se manifestam na forma de ruído na imagem, por serem imperfeições únicas de modelo e dispositivo, essas podem ser abordagens para identificação de fonte, [3] e [5].

Uma impressora a laser (electrophotographic printer) consiste em um rolo que é carregado eletrostaticamente e então um diodo envia um feixe de luz para retirar a essa carga em certos pontos do rolo, para poder ser formada a imagem. A tinta é então passada a esse rolo devido as forças eletrostáticas e finalmente a imagem é comprimida no papel. Imperfeições na velocidade angular das engrenagens da impressora cria um efeito banda, que se reproduz em linhas claras ou escuras perpendiculares ao processo de impressão [2], isso cria a assinatura única da impressora e pode ser explorado para identificação.

Uma impressora a jato de tinta usa rolos para puxar o papel e em cada passo o cartucho é passado em cima de uma linha no papel n vezes (conforme velocidade ou qualidade desejadas) e as gotas de tintas são atiradas no papel. Conforme a velocidade de impressão aumenta, a forma dos pontos variam, podendo ter também erro na posição dos pontos ou ainda a direção em que a gota foi atirada, além de vários outros fatores. Podemos citar [4] que faz uso de alguns desses fatores para identificação de impressoras.

Imperfeições sempre estão presentes nas impressoras, proporcionando assinaturas intrínsecas. Entretanto, estas nem sempre são suficientes, portanto são inseridos na impressão assinaturas extrínsecas no

documentos antes da impressão. Esses métodos devem persistir a impressão e digitalização para podermos identificar a fonte, entre eles podemos citar o espaçamento entre linhas, mudança de halftone ao longo da imagem, orientação do ponto em cada célula de halftone entre outros.

Podem ser adicionadas também características intrínsecas a impressora e, portanto, independente do documento a ser impresso. Podemos imprimir uma matriz de pontos amarelos, invisíveis aos olhos, ao longo do documento, controlar o tamanho do ponto em impressoras laser ou uma mescla de sinais senoidais nas linhas verticais.

Existem também técnicas contra-forenses: digitalização e impressão, que tenta fazer um exato da impressão. Esta não funciona perfeitamente, o usuário pode tentar remover os detalhes extrínsecos aos dispositivos, entretanto detalhes intrínsecos da impressora ou scanner são de difícil manipulação pois exigem conhecimento do hardware. Um efeito ruim da identificação da fonte de documentos pode também ocorrer, quando uma pessoa quer compartilhar documentos “perigosos” pode ser rastreada pela impressora que foi usada, nesse caso portanto devem ser excluídas as características intrínsecas e extrínsecas do dispositivo.

Em vista do crescimento da quantidade de folhas escritas fica óbvio a necessidade de um estudo para prevenir falsificação e duplicações desautorizadas. Para que esses problemas sejam evitados a computação forense analisa os processos físicos dos dispositivos junta com as implicações deles para efetivamente descobrir os dispositivos fontes. Portanto, essas pesquisas servem de grande contribuição para a sociedade.

3 Contribuições

Este artigo contribui para que pessoas com pouco ou nenhum conhecimento nessa área possam se informar sobre as inúmeras aplicações que existem para a computação, principalmente na área de impressoras e scanners. Os autores apresentam várias idéias comuns nessa área e exemplos de como aplicar essas técnicas nos vários problemas que afligem nossa sociedade, como pirataria e falsificação. Isso faz com que aquele que tiver interesse nessa área despertado nesse artigo, possa se aprofundar, pois o autor cita diversas referências com o detalhamento dos algoritmos citados.

4 Defeitos/Desvantagens

1. Os autores citam vários artigos para investigar a identificação de impressoras, mas não de scanners, ponto que também é muito relevante se quisermos identificar a origem de uma falsificação por exemplo.
2. Com a imensidão de técnicas existentes para identificação de dispositivos o autor cita apenas 4 ou 5 tipos diferentes. Para evitar que o artigo ficasse muito extenso, os autores podiam pelo menos mencionar onde poderíamos encontrar e ler sobre tais técnicas.
3. Os autores fazem referência aos métodos de identificação de impressoras, entretanto não deixam claro o propósito dos métodos de identificação extrínseca ou intrínseca. Um leitor deve subentender que tais técnicas seriam utilizadas em conjunto com os métodos de identificação de impressoras que não fazem de identificação por assinatura extrínseca ou intrínseca. Em suma, o texto poderia ser mais sucinto nesse assunto.

5 Trabalhos correlatos

5.1 Scanner identification using feature-based processing and analysis – Nitin Khanna et al. [5]

Relação com o artigo avaliado: Ambos tratam da computação forense na identificação da origem de scanners, entretanto esse artigo dá uma ênfase maior nesse ponto.

Descrição: Os autores descrevem uma técnica para identificação da origem do scanner através da análise do ruído característico de cada dispositivo. O algoritmo é diferente de uma análise comum de ruídos pois é levado em conta as colunas e linhas da imagem, sendo que as linhas são características de uma “foto” linear em um dado tempo t e as colunas são características do motor de passo e as imperfeições de movimento ligadas a ele.

Dependendo do tipo de compressão utilizada, os resultados são próximos a 100%, apesar de que mesmo com o uso de uma compressão JPEG em que há perdas de artefatos da foto, o método proposto pelos autores ainda consegue bons resultados. Esses resultados são inclusive muito superiores a métodos já propostos por outros autores.

5.2 Geometric distortion signatures for printer identification – Gaurav Sharma et al. [1]

Relação com o artigo avaliado: Ambos apresentam métodos para identificação de impressoras, entretanto o artigo mencionado dá um foco maior na identificação de impressoras a laser, explorando imperfeições geométricas produzidos por esses dispositivos.

Descrição: Esse artigo leva em conta as imperfeições na construção de impressoras a laser — principalmente por atrasos ou adiantamentos na velocidade angular dos rolos e imperfeições no movimento do polígono que reflete o laser no rolo — para determinar a impressora de origem. Os autores fazem uma análise geométrica do meio-tom produzido por diferentes impressoras. O processo consiste simplificada-mente em medir as distâncias reais entre os pontos em um computador, fazer a impressão, seguido de uma digitalização e medir novamente as distâncias entre os pontos. Da diferença dessas medidas temos uma assinatura da impressora de origem.

Os autores testam essa técnica em 6 impressoras diferentes e verificam a correlação das impressões das diferentes impressoras. Foram obtidas altas taxas de acerto para um threshold de 0,87. A grande vantagem desse método consiste na não necessidade de equipamentos especiais extrair informações da impressora, nem da necessidade de se conhecer especificidades de hardware de cada uma. A única requisição é que sejam feitas impressões para se obter uma média para análise.

6 Extensões

Os autores mencionam muito pouco a identificação de scanners, como esse também é um fator importante para identificação de falsificações e duplicações, isso também poderia ser mostrado no artigo. Como as várias técnicas de detecção de ruído, análise do motor de passo, imperfeições no scanner, entre vários outros pontos que podiam ser explorados.

7 Notas

1. Relevância: 9.0
2. Originalidade: 8.2
3. Qualidade científica: 8.5

4. Apresentação: 8.5

5. Nota final: 8.5

Referências

- [1] Bulan, Orhan, Junwen Mao e Gaurav Sharma: *Geometric distortion signatures for printer identification*. Acoustics, Speech, and Signal Processing, IEEE International Conference on, 0:1401–1404, 2009.
- [2] Chiang, Pei ju, Gazi N. Ali, Aravind K. Mikkilineni, Edward J. Delp, Jan P. Allebach e George T. c. Chiu: *Extrinsic signatures embedding using exposure modulation for information hiding and secure printing in electrophotographic devices*. In *Proceedings of the IS&T's NIP20: International Conference on Digital Printing Technologies*, páginas 295–300, 2004.
- [3] Gou, H., A. Swaminathan e M. Wu: *Robust scanner identification based on noise features*. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 6505 de *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, fevereiro 2007.
- [4] Kee, Eric e Hany Farid: *Printer profiling for forensics and ballistics*. In *MM&Sec '08: Proceedings of the 10th ACM workshop on Multimedia and security*, páginas 3–10, New York, NY, USA, 2008. ACM, ISBN 978-1-60558-058-6.
- [5] Khanna, N., A.K. Mikkilineni e E.J. Delp: *Scanner Identification Using Feature-Based Processing and Analysis*. Information Forensics and Security, IEEE Transactions on, 4(1):123 –139, mar. 2009, ISSN 1556-6013.