Identificação das teclas digitadas a partir da vibração mecânica

Gerson de Souza Faria e Hae Yong Kim

***Resumo*— Este artigo descreve um ataque que detecta as teclas pressionadas em um terminal de ponto de venda através da análise das vibrações mecânicas geradas quando as teclas são pressionadas. Usamos acelerômetros como sensores de vibração. O aparelho necessário para este ataque é de baixo custo e pode ser incorporado discretamente dentro do terminal. Obtivemos uma taxa de sucesso que varia de 58% a 82% em reconhecer as teclas pressionadas.**

Palavras-Chave—ataque não invasivo, segurança da informação, acelerômetros, vibração, senha, ponto de venda.

***Abstract*— This paper describes an attack that detects the sequence of keystrokes on a point of sale terminal through the analysis of mechanical vibrations generated when the keys are pressed. We use accelerometers as vibration sensors. The apparatus necessary for this attack is inexpensive and can be unobtrusively embedded within the terminal. We achieved a success rate ranging from 58% to 82% to recognize the keys.**

Keywords—side-channel attack, information security, accelerometer, vibration, password, POS.

I. INTRODUÇÃO

Atualmente, teclados mecânicos são a principal interface homem-máquina devido à sua facilidade de operação, eficiência e baixo custo. No mercado de meios de pagamento eletrônico, teclados são a escolha natural para a entrada de dados sigilosos, como senhas em terminais de ponto de venda, caixas eletrônicos etc. Na esfera governamental, os teclados mecânicos são usados para inserir os números de candidatos em urnas eletrônicas. Assim, a possibilidade de que alguém descubra a sequência de teclas digitadas (sem que o usuário perceba) é uma séria ameaça à segurança de sistemas.

O objetivo deste artigo é descrever um ataque físico não invasivo a terminais de ponto de venda (POS - *Point Of Sale*) que permite detectar as teclas digitadas a partir das vibrações mecânicas geradas no equipamento pelo ato de pressioná-las. Tais vibrações são capturadas por acelerômetros instalados no interior do terminal.

A possibilidade de um ataque similar foi sugerida por Kuhn [1], em que o autor sugere a possibilidade de identificar as teclas pressionadas pela análise vetorial das forças resultantes em determinados pontos do equipamento, por exemplo, nos suportes da base (pés). O autor afirma ser improvável que o roubo de informação por meio de canais secundários (*side-channels*) fique restrito aos domínios eletromagnético, óptico e acústico. Porém, segundo o nosso conhecimento, este ataque nunca foi testado experimentalmente. Nosso ataque não analisa as forças nas bases, mas sim as vibrações mecânicas geradas pelo ato de pressionar as teclas.

A. *Trabalhos relacionados*

Não encontramos na literatura ataques a teclados mecânicos pela análise de vibrações capturadas por meio de acelerômetros. Encontramos apenas ataques que analisam os sons gerados ao pressionar teclas de teclados de computador [2, 3, 4], teclas de terminais de caixas eletrônicos [2] e ataques acústicos em outros dispositivos, tais como impressoras matriciais [5]. Shamir e Tromer apresentam em [14] uma prova de conceito de criptanálise baseada em emanações acústicas de computadores pessoais. Cai e Chen apresentam um ataque que permite inferir os dígitos pressionados no teclado virtual de celulares baseados no sistema Android analisando o movimento do aparelho capturado pelo acelerômetro interno do mesmo [11]. Uma análise geral sobre a vulnerabilidade de sensores de celulares é apresentada em [16].

B. *Contribuição*

A principal contribuição deste artigo é expor uma vulnerabilidade na arquitetura de terminais POS. Apresentamos elementos suficientes que mostram que ataques não invasivos a terminais POS para roubo de senha são possíveis e podem ser efetuados a um baixíssimo custo.

II. DESCRIÇÃO DAS VULNERABILIDADES IDENTIFICADAS

Equipamentos como os terminais POS possuem mecanismos de detecção de violação física

(*tampering*), de modo a auto-destruir informação sensível, como chaves criptográficas contidas em seu perímetro de segurança em caso de detecção de violação. No entanto, uma inspeção visual mostrou que vários terminais POS possuem em sua parte inferior uma tampa removível de serviço e manutenção de modo a oferecer acesso legítimo aos conectores de cartões SAM (*Security Authentication Module*), responsáveis pela segurança da comunicação do sistema bem como pela autenticação com as redes de serviços. Tal espaço possibilita a implantação de dispositivos de coleta ilegal de informação ("*bugs*", que em nosso caso são os acelerômetros). Um mecanismo de detecção de violação comumente utilizado são selos do tipo "*void seal*" (Figura 1), aplicados no encontro da tampa com o corpo do terminal, indicando visualmente a prévia abertura do compartimento. Obviamente, o consumidor não costuma prestar atenção a tal item no momento de digitar a sua senha.

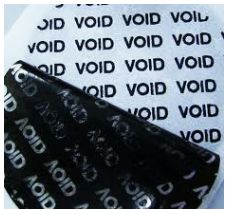

Fig. 1. Exemplo de selo de detecção de violação utilizado para evidenciar a abertura de equipamentos.

Além do acesso ao compartimento, outra condição que beneficiaria ainda mais um ataque é a alimentação elétrica disponível nos terminais dos conectores SAM, que poderia ser utilizada para alimentar cartões SAM falsos contendo acelerômetros e possíveis circuitos auxiliares de comunicação. Como consequência, o ataque poderia se tornar não invasivo e não detectável no terminal, sem uso de fios e baterias aparentes.

III. MONTAGEM DO ATAQUE

A. *Posicionamento dos sensores*

A abordagem aqui adotada pode ser aplicada em praticamente qualquer terminal POS que possua teclado mecânico em seu corpo. A matriz de teclas do equipamento utilizado é padrão, como pode ser observada na Figura 2.

Fig. 2. Formato da matriz de teclas do equipamento utilizado no ataque.

Dois acelerômetros são utilizados no ataque. As placas dos mesmos foram envolvidas em um pedaço de espaguete termo-retrátil e coladas nas posições da Figura 3. Os acelerômetros não possuem qualquer conexão elétrica com o terminal, sendo que no protótipo foram utilizados cabos de conexão ligando os sensores ao sistema de aquisição. Após acondicionados os sensores e os cabos, a tampa original foi recolocada no terminal.

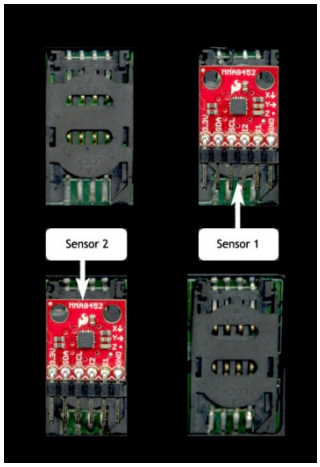

Fig. 3. Vista da parte inferior do terminal, sem a tampa de acesso aos conectores SAM, com a disposição dos acelerômetros utilizados. A máscara preta na imagem foi aplicada para não expor detalhes que pudessem identificar o modelo do equipamento.

B. *Acelerômetros e sistema de aquisição*

Os acelerômetros utilizados são MMA8452 de 12 bits e três eixos, da família Freescale Xtrinsic [6]. O sistema de aquisição foi desenvolvido na plataforma Arduino Mega 2560 [12], como mostra a Figura 4.

A escala de aceleração adequada foi ±2g e a taxa de amostragem adotada foi a máxima para o acelerômetro, 800 amostras/s. Utilizamos as informações de aceleração nos três eixos. Foram coletadas 300 amostras para cada pressionamento de tecla, para os três eixos dos dois sensores, totalizando 1800 elementos por teclagem.

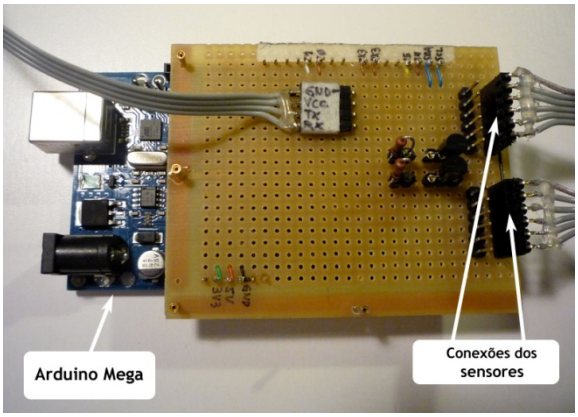

Fig. 4. O hardware do sistema de aquisição de dados consta de um Arduino Mega 2560 e uma placa de montagem sobre este, onde são conectados os cabos dos sensores.

C. *Obtenção das amostras*

De modo a capturar diferentes formas de teclar cada uma das teclas, o processo de obtenção das amostras envolveu 2 pessoas em 5 sessões de teclagens. Cada sessão foi composta pelo pressionamento de 40 vezes cada uma das 12 teclas da matriz 4x3 da Figura 2. Um dos testadores executou três sessões e o outro duas sessões. No total, adquirimos 2400 teclagens, 200 para cada tecla. Durante o processo de coleta das amostras, notamos que cada pessoa possui uma maneira distinta de teclar, envolvendo variações na intensidade, na posição e na permanência do dedo sobre a tecla. Desenvolvemos um programa MATLAB para a leitura das amostras do sistema de aquisição. Em todas as sessões, o terminal ficou em repouso sobre uma mesa, tendo um *mousepad* como base. As amostras foram tomadas com o terminal desligado.

IV. A ABORDAGEM ADOTADA

A. *Propriedades do sistema em análise*

Na teoria clássica de sistemas, a resposta do sistema a uma entrada impulsiva é a sua função de transferência. Se o sistema for linear, a sua saída é dada pela convolução entre a resposta impulsiva e o sinal de entrada [7]. Contudo, o sistema em análise é mecanicamente complexo, havendo acoplamento entre componentes fixos e móveis, amortecedores nas teclas, nos pés etc. Diante disto, não é razoável supor uma hipótese de linearidade para o sistema. Além disso, as pessoas podem apertar uma tecla de várias formas distintas, com maior ou menor intensidade, com maior ou menor permanência da pressão na tecla, com variação nas componentes de força etc. (Figura 5). Dessa forma, não adotamos a abordagem de identificação de sistemas, mas sim, abordamos o problema como um caso de classificação de padrões.

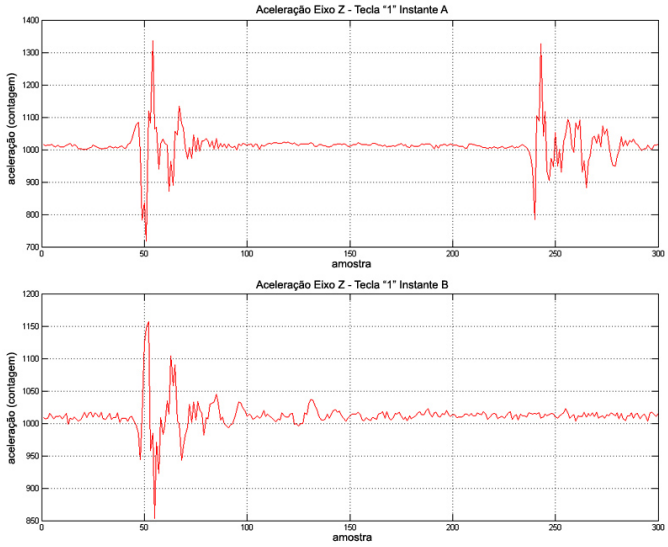

Fig. 5. Exemplos de duas amostras da aceleração do eixo Z (normal ao terminal) do mesmo sensor, da mesma tecla, efetuadas pela mesma pessoa. Os dois gráficos são bastante diferentes: o sentido da aceleração está invertido; o amortecimento e a amplitude são distintos. Além disso, o gráfico superior mostra a vibração de soltura da tecla (o segundo pulso). Este evento não ocorre no gráfico inferior, pois ocorreu fora do intervalo de aquisição.

B. *Características utilizadas*

Cada amostra é representada por um vetor de dimensão 6:

$$\mathbf{v} = (v_x, v_y, v_z, w_x, w_y, w_z)$$

sendo:

- v_x, v_y, v_z = vetores coluna com 300 amostras do sensor 1, eixos *x, y, z*.
- w_x, w_y, w_z = vetores coluna com 300 amostras do sensor 2, eixos *x, y, z*.

O vetor **v** também pode ser considerado como uma matriz 300×6.

As características (*features*) que utilizamos procuram identificar as dependências mútuas entre os sinais do vetor **v** para reconhecer as teclas. Deste modo, não consideramos representações isoladas dos sinais (e.g. espectro de frequência, coeficientes de auto regressão etc.). As seguintes características foram testadas (rótulos adotados em negrito):

- **xcorr**: correlação cruzada entre os sinais de mesma coordenada, i.e., correlação entre (v_x, w_x), (v_y, w_y) e (v_z, w_z), com m *lags,* centrados em zero, para m = ±5. A dimensão do vetor de características resultante é $3\times(2|m| + 1) = 33$.

- **pca**: vetores normalizados das p componentes principais da *Principal Component Analysis* (PCA) [8, cap.6] de **v,** para p = 3. Dimensão do vetor de características é 6p = 18.
- **cov**: matriz de covariâncias do vetor **v**, dada por:

  $$\mathbf{C_v} = (\mathbf{v} - \boldsymbol{\mu}_\mathbf{v})^T (\mathbf{v} - \boldsymbol{\mu}_\mathbf{v})$$

  Sendo $\boldsymbol{\mu}_\mathbf{v}$ o vetor de médias das colunas de **v**. A matriz de covariâncias resultante é uma matriz simétrica 6×6. Consideramos apenas o triângulo superior dessa matriz, resultando um vetor de características de dimensão 21.
- **white**: a matriz de *whitening* (ou *sphering*) de um vetor aleatório de média nula $\mathbf{z} = (z_1 \ldots z_n)$ é uma transformação linear calculada sobre **z** de forma a tornar seus elementos $\mathbf{z}_i$ decorrelacionados e com matriz de covariância **I** (matriz identidade). [8, p.140].

  Sejam:

  $\mathbf{E} = (\mathbf{e_1} \ldots \mathbf{e_n})$ a matriz cujas colunas são os autovetores de norma unitária da matriz de covariância $\mathbf{C_z}$;

  $\mathbf{D} = \text{diag}\ (\mathbf{d_1} \ldots \mathbf{d_n})$ a matriz diagonal dos autovalores da matriz de covariância $\mathbf{C_z}$.

  Assim, a matriz de *whitening* é dada por:

  $$\mathbf{W} = \mathbf{D}^{-1/2}\mathbf{E}^T$$

  O vetor transformado, de média nula e matriz de covariância **I** é então:

  $$\mathbf{y} = \mathbf{W}\mathbf{z}.$$

  A matriz de transformação **W** é utilizada como característica, possuindo dimensão = n^2, onde n é a dimensão do vetor **z**. Para a característica rotulada de **white**, utilizamos a matriz **W** calculada para os pares (v_x, w_x), (v_y, w_y) e (v_z, w_z), resultando num vetor de características de dimensão $3 \times 2^2 = 12$.
- **whitecruz**: matriz de *whitening* das combinações das 6 componentes do vetor **v** tomadas 2 a 2, resultando num vetor de características de dimensão $15 \times 2^2 = 60$.

Desenvolvemos um programa MATLAB para a geração das características. Para mais informações sobre o cálculo das mesmas, indicamos [8, 15].

C. Classificadores

Para cada conjunto de características do item anterior, testamos os seguintes classificadores (rótulos em negrito):

- **MLP**: rede neural artificial tipo perceptron multicamadas, na configuração fixa entrada-30-30-N, sendo N o número de classes;
- **RTree**: árvores aleatórias;
- **SVM**: máquina de vetores de suporte (apenas o kernel linear foi utilizado).

Utilizamos a implementação dos classificadores da biblioteca OpenCV 2.3 [17]. Todos os classificadores operaram no modo de múltiplas classes. Do total de amostras utilizadas, 80% foram destinadas a treino e 20% a teste, escolhidas de forma aleatória. Para mais informações sobre os classificadores, indicamos [9, 10, 13].

D. Esquemas de classificação dos sinais

O conjunto de dados de treino foi composto a partir de três esquemas de classes. Foram adotadas classes de 'linhas', 'colunas' e 'teclas'. No primeiro e segundo casos, as amostras das teclas foram agrupadas em sua linha/coluna. Para as classes de 'teclas' não há agrupamento.

V. RESULTADOS E DISCUSSÃO

As cinco melhores combinações de características e classificadores podem ser vistas no gráfico da Figura 6, com os três esquemas de classificação: 'linhas', 'colunas' e 'teclas'. O produto das probabilidades de acerto de linhas e colunas é designado no gráfico como 'colunas x linhas'. Observamos que a taxa de acerto de colunas é extremamente elevada e bem superior à taxa de acerto de linhas. A taxa de acerto da classificação 'teclas' é muito próxima ao produto 'colunas x linhas', indicando que o reconhecimento das teclas está limitado pela capacidade de reconhecimento das linhas. Deste modo, a taxa de reconhecimento da melhor combinação (perceptron multicamadas + matriz de covariâncias) ficou em torno de 68% com desvio padrão $\sigma \sim 5\%$. Tabelas I e II mostram as matrizes confusão para o melhor caso.

Notamos que é mais fácil discriminar as colunas do que as linhas, pois a variação da amplitude da vibração no sentido transversal é maior do que na longitudinal, devido à geometria do terminal: o comprimento do terminal é aproximadamente o dobro de sua largura e as linhas estão mais próximas entre si do que as colunas.

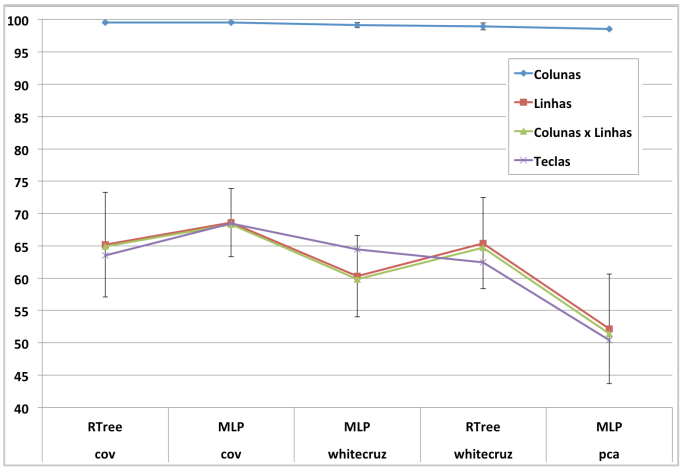

Fig. 6. Taxas de acertos (eixo vertical) obtidos para as cinco melhores combinações de métodos de classificação e características (eixo horizontal). O reconhecimento de colunas é muito superior ao de linhas.

TABELA I. MATRIZ CONFUSÃO DE LINHAS

	1	2	3	4	acerto (%)
1	**95**	21	0	0	81,9
2	24	**80**	19	7	61,5
3	3	20	**70**	26	58,8
4	0	11	21	**83**	72,2

TABELA II. MATRIZ CONFUSÃO DE COLUNAS

	1	2	3	acerto (%)
1	**162**	1	0	99,4
2	0	**159**	0	100,0
3	0	1	**157**	99,4

VI. TRABALHOS FUTUROS

Em todos os nossos testes, utilizamos apenas um tipo de características por vez. Talvez as taxas de reconhecimento possam ser melhoradas alimentando o classificador com uma combinação de vários tipos de características ao mesmo tempo. Pensamos ainda que uma modelagem das dependências entre os eixos de aceleração poderia auxiliar no posicionamento dos acelerômetros e melhorar os resultados. Dado que parte dos usuários opera o terminal segurando-o com uma das mãos, seria importante testar o ataque proposto para este caso.

VII. CONCLUSÕES

A análise das vibrações mecânicas de um terminal POS decorrente do ato de teclar permitiu descobrir os dígitos teclados com taxas de acerto entre 58% e 82% por dígito. O reconhecimento das teclas ficou limitado somente pela capacidade de discriminar linhas, uma vez que o acerto no reconhecimento de colunas foi de 99,6%. A possibilidade de acesso físico ao compartimento de conectores SAM facilita (embora não seja necessária) a instalação discreta de acelerômetros e, eventualmente, de sistemas de comunicação. Isto aumentaria a vulnerabilidade do equipamento.

REFERÊNCIAS

[1] M. G. Kuhn. *Compromising emanations: eavesdropping risks of computer displays* - Technical Report UCAM-CL-TR-577, University of Cambridge, Computer Laboratory, Dec. 2003, p.132.

[2] D. Asonov, R. Agrawal. *Keyboard Acoustic Emanations* – Proceedings IEEE Symposium on Security and Privacy, May 2004.

[3] L. Zhuang et al. *Keyboard Acoustic Emanations Revisited* - Proceedings of the 12th ACM Conference on Computer and Communications Security, Nov. 2005, pp. 373-382

[4] Y. Berger et al. *Dictionary attacks using keyboard acoustic emanations*, Proceedings of the 13th ACM conference on Computer and communications security, 2006, Alexandria, Virginia, USA.

[5] M. Backes. *Acoustic side-channel attack on printers* - USENIX Security'10 Proceedings of the 19th USENIX conference on Security, 2010.

[6] Freescale. MMA 8452: Xtrinsic 3-Axis, 12 bit Accelerometer (http://www.freescale.com/webapp/sps/site/prod_summary.jsp?code=MMA8452Q&fr=g) Acessado em Janeiro 2012.

[7] Oppenheim, Alan V.; Schafer, R. W.; and Buck, J. R. (1999). *Discrete-time signal processing*. Upper Saddle River, N.J.: Prentice Hall

[8] Hyvärinen, Aapo & Karhunen, Juha & Oja, Erkki. *Independent Component Analysis*. John Wiley & Sons, 2001.

[9] C.J.C Burges. *A Tutorial on Support Vector Machines for Pattern Recognition,* 1998 - (http://www.umiacs.umd.edu/%7Ejoseph/support-vector-machines4.pdf). Acessado em Fevereiro 2012.

[10] G. Bradski, A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly, 2008

[11] L. Cai and H. Chen. 2011. *TouchLogger: inferring keystrokes on touch screen from smartphone motion.* In *Proceedings of the 6th USENIX conference on Hot topics in security* (HotSec'11). USENIX Association, Berkeley, CA, USA, 9-9.

[12] Arduino homepage: http://www.arduino.cc/. Acessado em Fevereiro 2012.

[13] S. Haykin, *Neural Networks and Learning Machines – Third Edition* – Pearson. 2009.

[14] A. Shamir, E. Tromer. *Acoustic cryptanalysis - On nosy people and noisy machines* (http://www.cs.tau.ac.il/~tromer/acoustic/). Acessado em Dezembro 2012.

[15] D. H. Kil, F. B. Shin. *Pattern Recognition and Prediction with applications to signal characterization*. American Institute of Physics, Woodbury, New York, 1996.

[16] L. Cai, S. Machiraju, H. Chen. *Defending against sensor-sniffing attacks on mobile phones*. In *Proceedings of the 1st ACM workshop on Networking, systems, and applications for mobile handhelds* (MobiHeld '09). ACM, New York, NY, USA, 31-36. 2009

[17] OpenCV homepage: http://opencv.willowgarage.com/wiki. Acessado em Fevereiro 2012.

Gerson de Souza Faria e Hae Yong Kim: Escola Politécnica, Universidade de São Paulo - São Paulo, Brasil. emails: gerson.faria@usp.br, hae@lps.usp.br.
Este trabalho foi financiado parcialmente pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes).