References

[1] A. Heuser and M. Zohner, "Intelligent machine homicide," in *Constructive Side-Channel Analysis and Secure Design* (W. Schindler and S. A. Huss, eds.), (Berlin, Heidelberg), pp. 249–264, Springer Berlin Heidelberg, 2012.

[2] T. Bartkewitz and K. Lemke-Rust, "Efficient template attacks based on probabilistic multi-class support vector machines," in *Smart Card Research and Advanced Applications* (S. Mangard, ed.), (Berlin, Heidelberg), pp. 263–276, Springer Berlin Heidelberg, 2013.

[3] G. Hospodar, B. Gierlichs, E. De Mulder, I. Verbauwhede, and J. Vandewalle, "Machine learning in side-channel analysis: a first study," *Journal of Cryptographic Engineering*, vol. 1, p. 293, Oct 2011.

[4] L. Lerman, G. Bontempi, and O. Markowitch, "Power analysis attack: An approach based on machine learning," vol. 3, p. ied Cryptography, 01 2014.

[5] L. Lerman, G. Bontempi, and O. Markowitch, "A machine learning approach against a masked AES," *Journal of Cryptographic Engineering*, vol. 5, pp. 123–139, Jun 2015.

[6] R. Gilmore, N. Hanley, and M. O'Neill, "Neural network based attack on a masked implementation of AES," in *2015 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*, pp. 106–111, May 2015.

[7] H. Maghrebi, T. Portigliatti, and E. Prouff, "Breaking cryptographic implementations using deep learning techniques," in *Security, Privacy, and Applied Cryptography Engineering* (C. Carlet, M. A. Hasan, and V. Saraswat, eds.), vol. 10076, pp. 3–26, Springer International Publishing, 2016.

[8] M. Brisfors and S. Forsmark, "DLSCA: a Tool for Deep Learning Side Channel Analysis," p. 6, 2019.

[9] Fanliang Hu, Huanyu Wang, and Junnian Wang, "Cross-subkey deep-learning side-channel analysis.," *IACR Cryptol. ePrint Arch.*, 2021.

[10] Huanyu Wang and Elena Dubrova, "Federated learning in side-channel analysis.," *IACR Cryptol. ePrint Arch.*, 2020.

[11] Huanyu Wang and Elena Dubrova, "Tandem deep learning side-channel attack against fpga implementation of aes.," *IACR Cryptol. ePrint Arch.*, 2020.

[12] Kalle Ngo, Elena Dubrova, Qian Guo 001, and Thomas Johansson 001, "A side-channel attack on a masked ind-cca secure saber kem implementation.," *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2021.

[13] Yang Yu, Michail Moraitis, and Elena Dubrova, "Can deep learning break a true random number generator?," *IEEE Trans. Circuits Syst. II Express Briefs*, 2021.

[14] D. Das, A. Golder, J. Danial, S. Ghosh, A. Raychowdhury, and S. Sen, "X-deepsca: Cross-device deep learning side channel attack," pp. 1–6, 06 2019.

[15] S. Bhasin, A. Chattopadhyay, A. Heuser, D. Jap, S. Picek, and R. R. Shrivastwa, "Mind the portability: A warriors guide through realistic profiled side-channel analysis." Cryptology ePrint Archive, Report 2019/661, 2019. `https://ia.cr/2019/661`.

[16] R. Benadjila, E. Prouff, R. Strullu, E. Cagli, and C. Dumas, "Study of deep learning techniques for side-channel analysis and introduction to ASCAD database," p. 45, 2018.

[17] H. Wang, M. Brisfors, S. Forsmark, and E. Dubrova, "How Diversity Affects Deep-Learning Side-Channel Attacks," in *2019 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC)*, (Helsinki, Finland), pp. 1–7, IEEE, Oct. 2019.

[18] M. Brisfors and S. Forsmark, "Deep-learning side-channel attacks on aes," 2019.

[19] M. Brisfors, S. Forsmark, and E. Dubrova, "How Deep Learning Helps Compromising USIM," in *Smart Card Research and Advanced Applications* (P.-Y. Liardet and N. Mentens, eds.), vol. 12609, pp. 135–150, Cham: Springer International Publishing, 2021. Series Title: Lecture Notes in Computer Science.

[20] C. Devine, "A Practical Guide to Differential Power Analysis of USIM cards," p. 13.

[21] R. Wang, H. Wang, and E. Dubrova, "Far Field EM Side-Channel Attack on AES Using Deep Learning," in *Proceedings of the 4th ACM Workshop on Attacks and Solutions in Hardware Security*, (Virtual Event USA), pp. 35–44, ACM, Nov. 2020.

[22] H. Wang, R. Wang, M. Brisfors, and E. Dubrova, "Advanced far field em side-channel attack on aes based on deep learning," 2021.

[23] N. Parmar, A. Vaswani, J. Uszkoreit, Łukasz Kaiser, N. Shazeer, A. Ku, and D. Tran, "Image transformer," 2018.

[24] D. Grechishnikova, "Transformer neural network for protein-specific de novo drug generation as a machine translation problem," *Scientific Reports*, vol. 11, p. 321, Jan. 2021.

[25] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210–229, 1959.

[26] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain.," *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958.

[27] K. D. Foote, "A Brief History of Machine Learning," Mar. 2019. `https://www.dataversity.net/a-brief-history-of-machine-learning/`.

[28] "UCI Machine Learning Repository: Iris Data Set." `https://archive.ics.uci.edu/ml/datasets/iris`.

[29] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986.

[30] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017.

[32] H. Ramsauer, B. Schäfl, J. Lehner, P. Seidl, M. Widrich, L. Gruber, M. Holzleitner, M. Pavlovic, G. K. Sandve, V. Greiff, D. P. Kreil, M. Kopp, G. Klambauer, J. Brandstetter, and S. Hochreiter, "Hopfield networks is all you need," *CoRR*, vol. abs/2008.02217, 2020.

[33] D. Kahn, *The Codebreakers: The comprehensive history of secret communication from ancient times to the internet.* Simon and Schuster, 1996.

[34] J. Daemen and V. Rijmen, "The rijndael block cipher," p. 45, March 1999.

[35] P. C. Kocher, "Timing attacks on implementations of di e-hellman, RSA, DSS, and other systems," p. 10, 1996.

[36] P. Kocher, J. Jaffe, and B. Jun, "Differential Power Analysis," in *Advances in Cryptology — CRYPTO' 99* (M. Wiener, ed.), (Berlin, Heidelberg), pp. 388–397, Springer Berlin Heidelberg, 1999.

[37] G. Camurati, S. Poeplau, M. Muench, T. Hayes, and A. Francillon, "Screaming Channels: When Electromagnetic Side Channels Meet Radio Transceivers," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, (Toronto Canada), pp. 163–177, ACM, Oct. 2018.

[38] Sysmocom, "Sysmocom store - sysmoUSIM-SJS1 SIM + USIM." `http://shop.sysmocom.de/products/sysmousim-sjs1`.

[39] NewAE Technology Inc., "ChipWhisperer." `https://newae.com/tools/chipwhisperer/`.

[40] R. Benadjila, M. Renard, D. Elbaze, and P. Trébuchet, "LEIA: the Lab Embedded ISO7816 Analyzer A Custom Smartcard Reader for the ChipWhisperer," p. 30, 2019.

[41] William Falcon, "Pytorch lightning," March 2019. https://www.pytorchlightning.ai/.

[42] Keras, "Keras documentation," May 2019. https://keras.io/.

[43] "Pearson's Correlation Coefficient," in *Encyclopedia of Public Health* (W. Kirch, ed.), pp. 1090–1091, Dordrecht: Springer Netherlands, 2008.

[44] Z. Lu and K.-H. Yuan, *Welch's t test*, pp. 1620–1623. 01 2010.

[45] Keras Team, "Keras documentation: Text classification with Transformer," May 2020. https://keras.io/examples/nlp/text_classification_with_transformer/.

[46] L. Liu, Z. Qu, Z. Chen, Y. Ding, and Y. Xie, "Transformer acceleration with dynamic sparse attention," 2021.

[47] R. Child, S. Gray, A. Radford, and I. Sutskever, "Generating long sequences with sparse transformers," 2019.

[48] Google AI, "Google AI Blog: Constructing Transformers For Longer Sequences with Sparse Attention Methods," March 2021. https://ai.googleblog.com/2021/03/constructing-transformers-for-longer.html.