# Intelligent malware detection and classification (Intelligent Malware Sandbox)

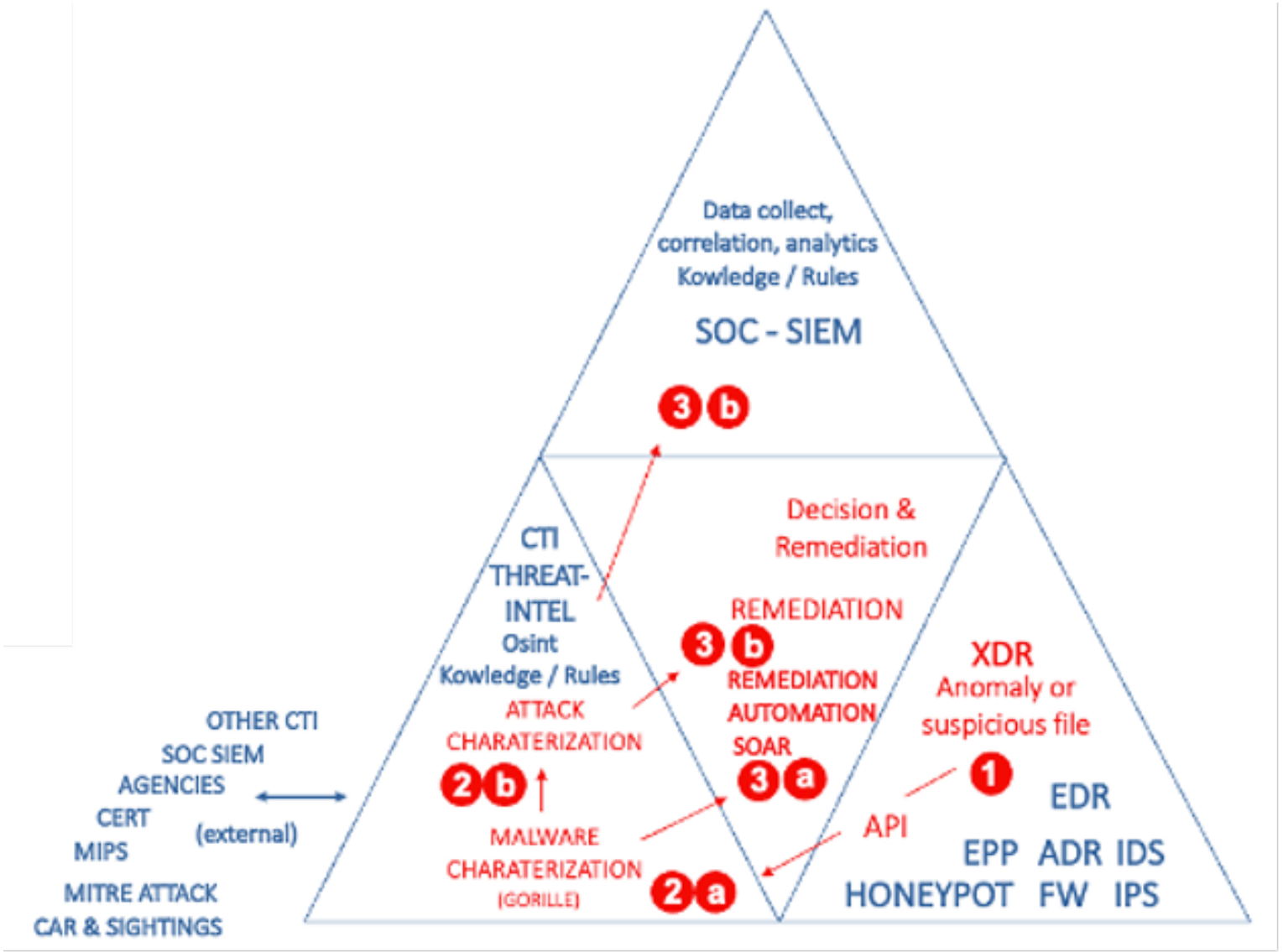
Dr. Sanjeev Kumar

Scientist E, C-DAC Mohali

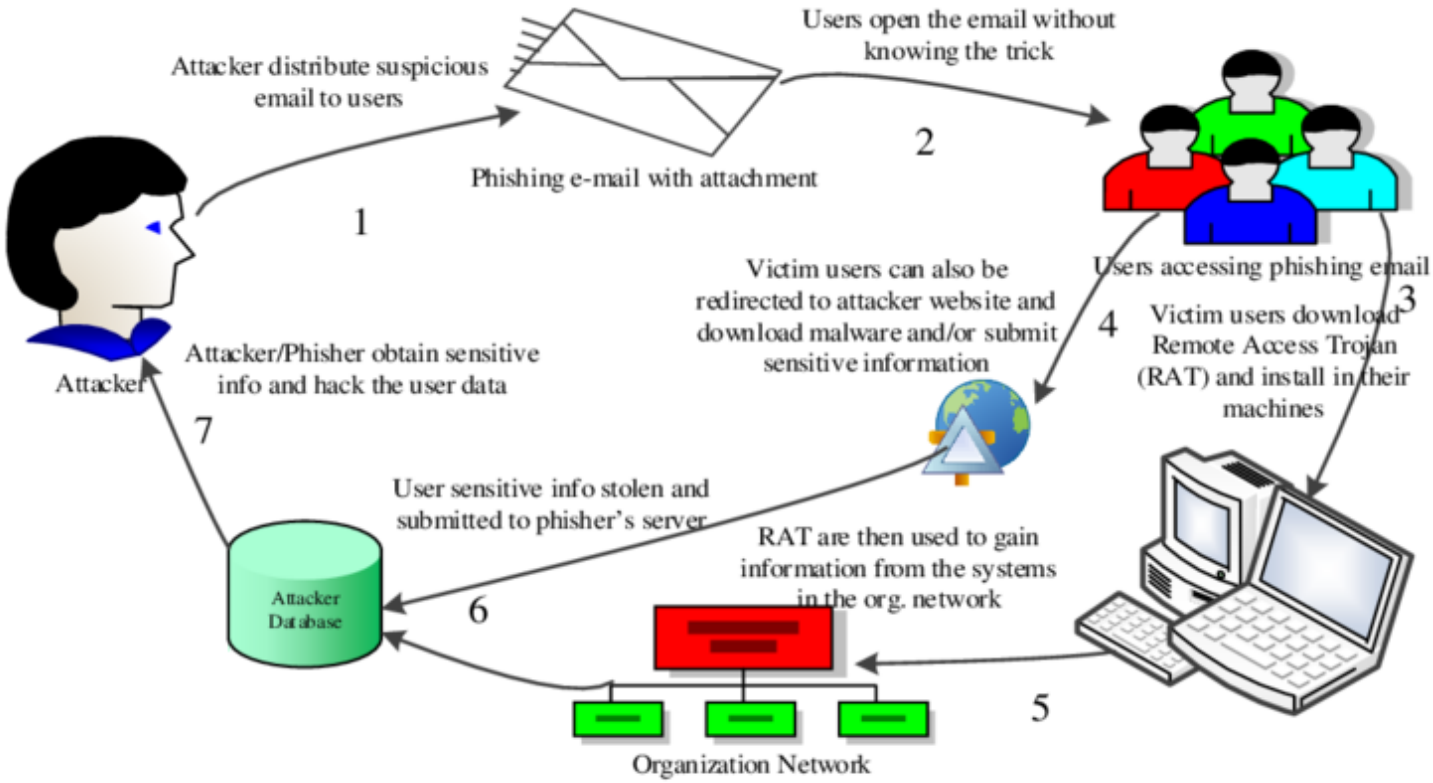
January 1, 2025

# Table of Contents I

# Security Operation Workflow



# Malware based Phishing Attacks




**Reference:** [KM22]

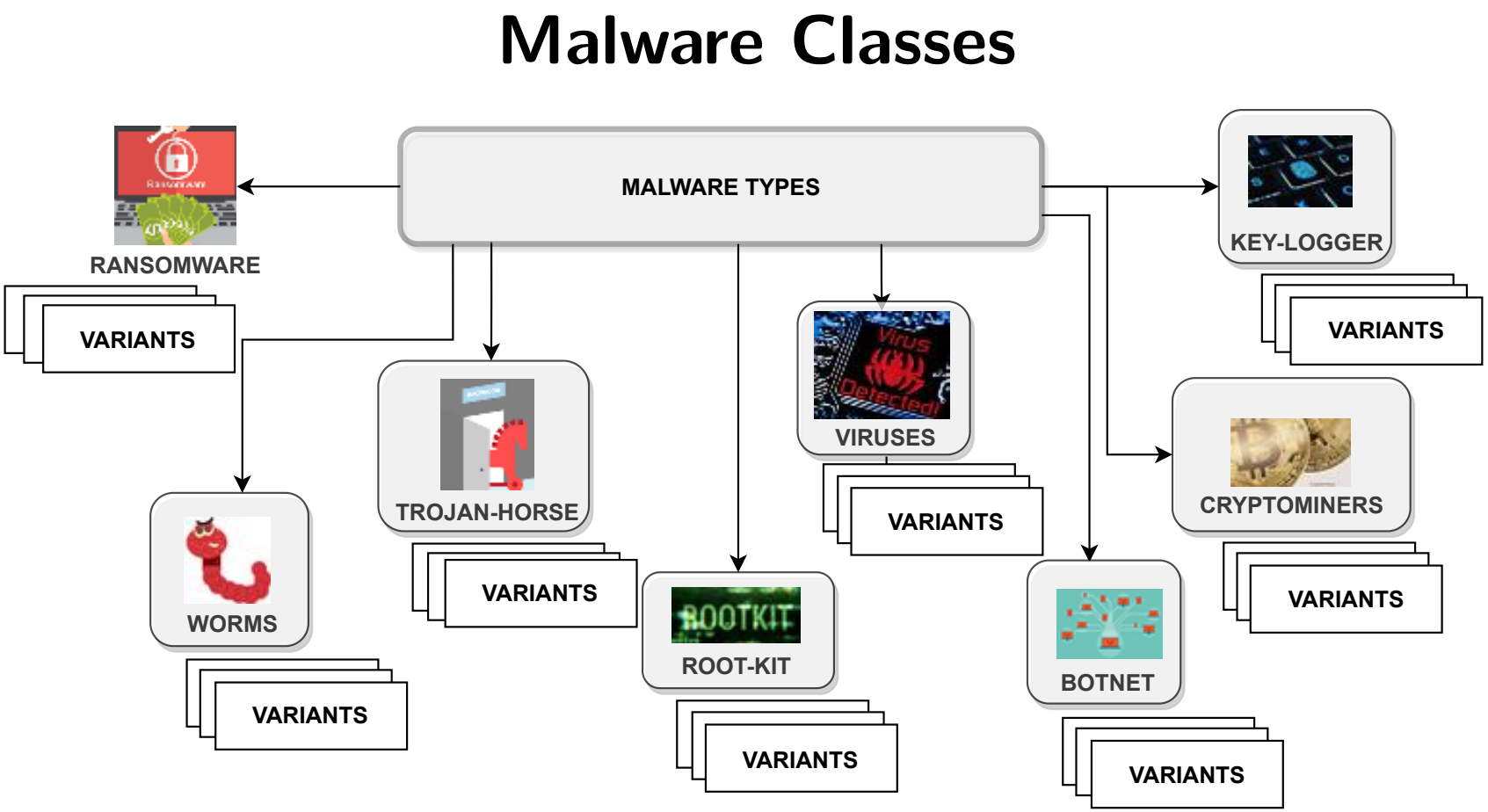
# What is Malware

- Malware, short for malicious software, refers to any intrusive software developed by cybercriminals (often called hackers) to steal data and damage or destroy computers and computer systems.
- Malware is malicious software and refers to any software that is designed to cause harm to computer systems, networks, or users.
- Malware includes various types of cyber threats such as Viruses, worms, Trojan viruses, spyware, adware, botnets, advanced persistent threats (APT), ransomware etc.
- Malware can take many forms. Individuals and organizations need to be aware of the different types of malware and take steps to protect their systems, such as using antivirus software, keeping software and systems up-to-date, and being cautious when opening email attachments or downloading software from the internet.

# Key Trends of Malware Attacks

- Windows container malware
- Mobile Malware/Industrial IoT/ Linux
- Cryptojacking
- Malware disruption
- Malware as a business

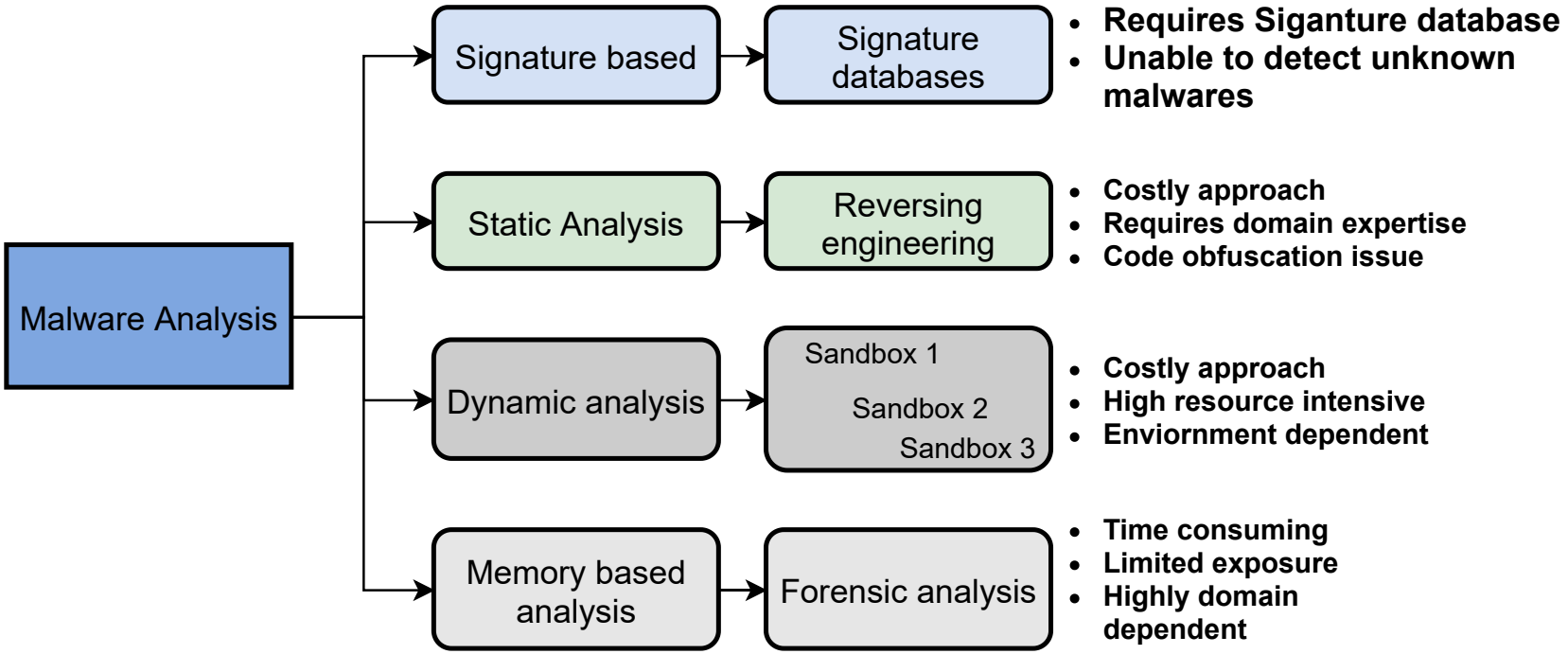
# Different Malware Threats



# What is Malware Analysis

- Malware analysis is the process of studying harmful software, or malware, to understand its characteristics, origins, and potential effects.
- The goal of malware analysis is to help security teams make informed decisions about how to detect and respond to malware.
- Malware analysis is the study of the unique features, objectives, sources, and potential effects of harmful software and code, such as spyware, viruses, malvertising, and ransomware. It analyzes malware code to understand how it varies from other kinds

# Malware Analysis Techniques

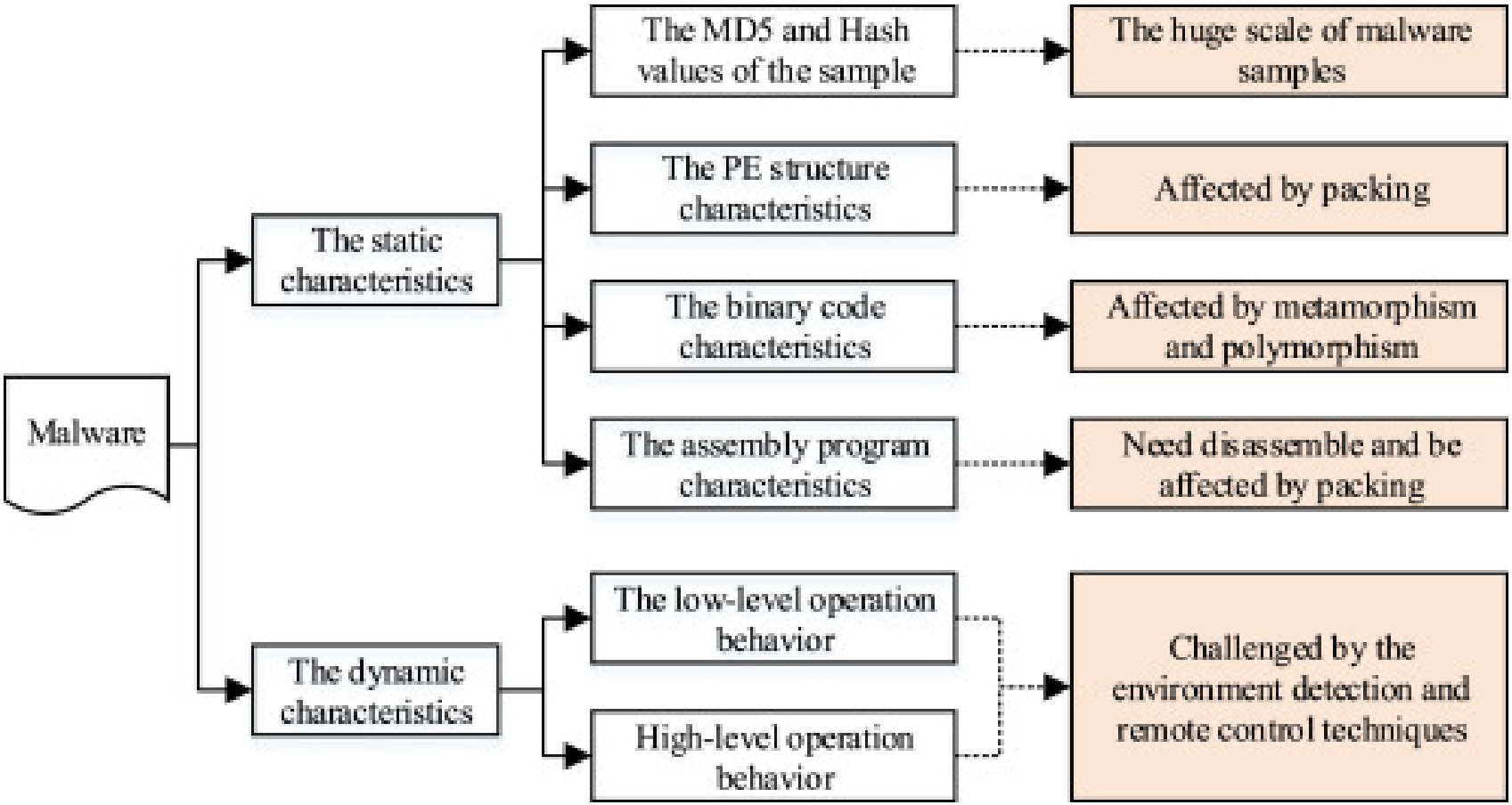



**Feature extraction through these approaches are highly dependence on domain knowledge**

# What are benefits of Malware Analysis

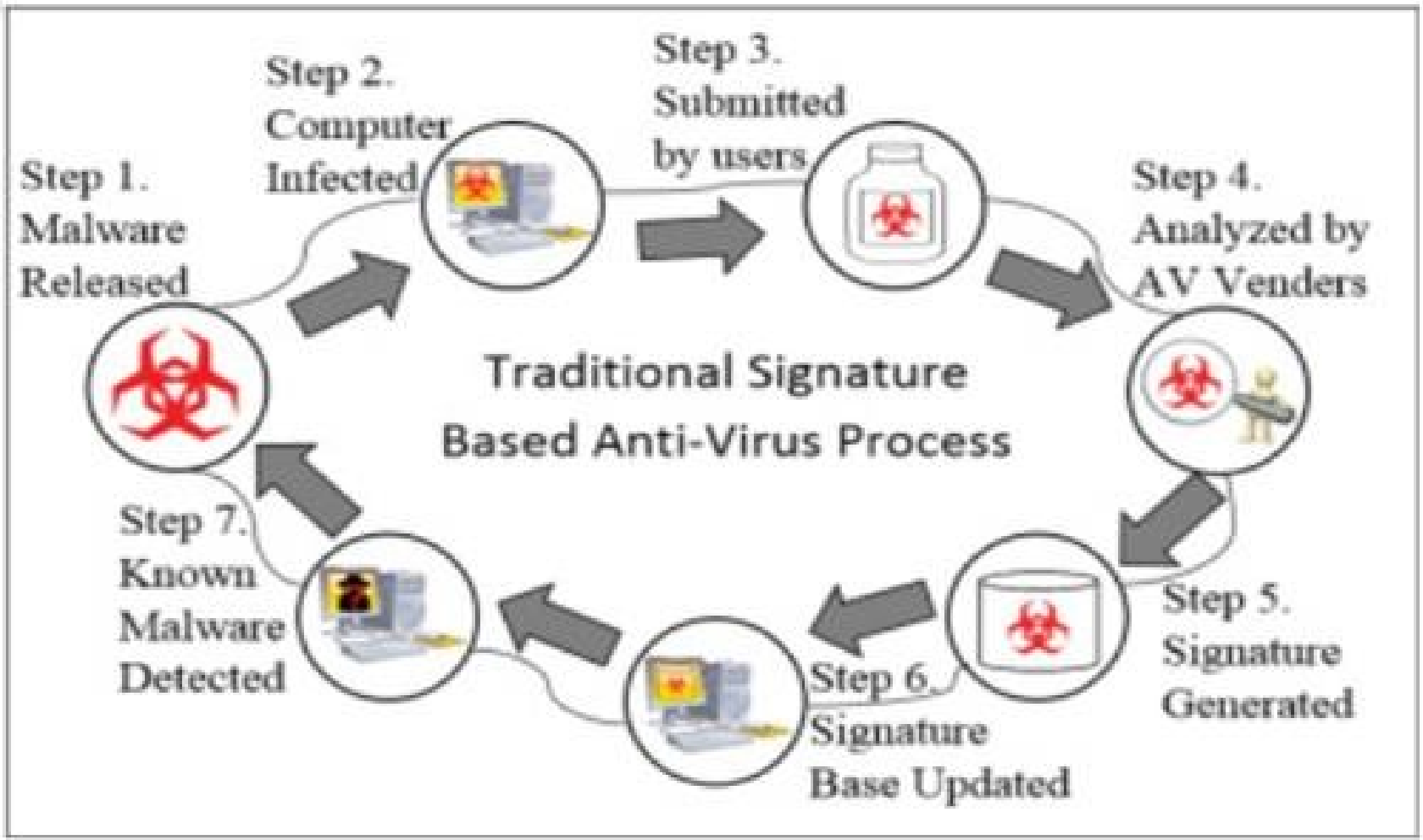
- Figure out how much damage an intrusion caused.
- Identify who may have installed malware inside the system.
- Determine the attack's level of sophistication.
- Pinpoint the exact vulnerability the malware exploited to access your system.

# Malware Analysis Techniques cont.

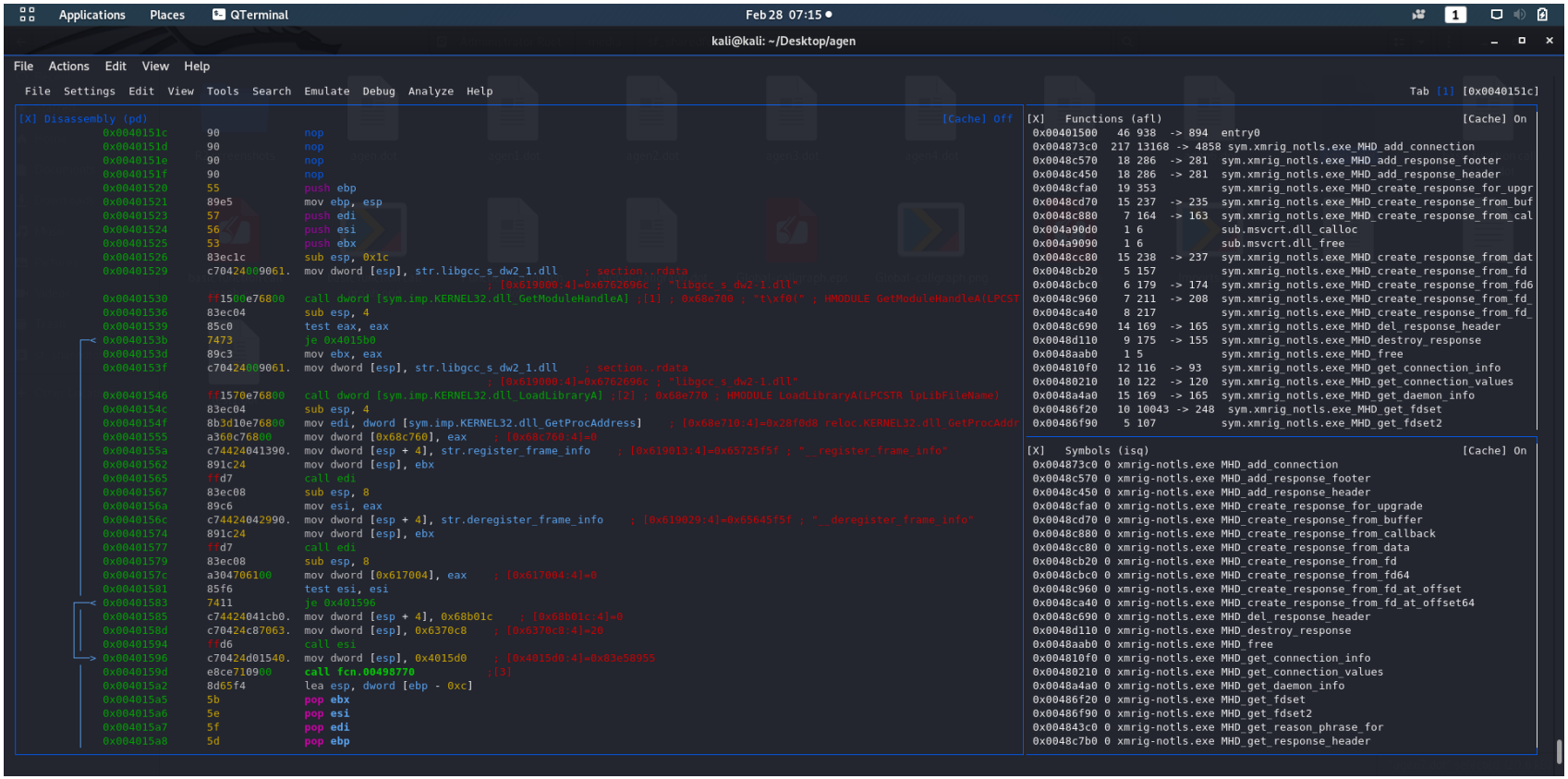


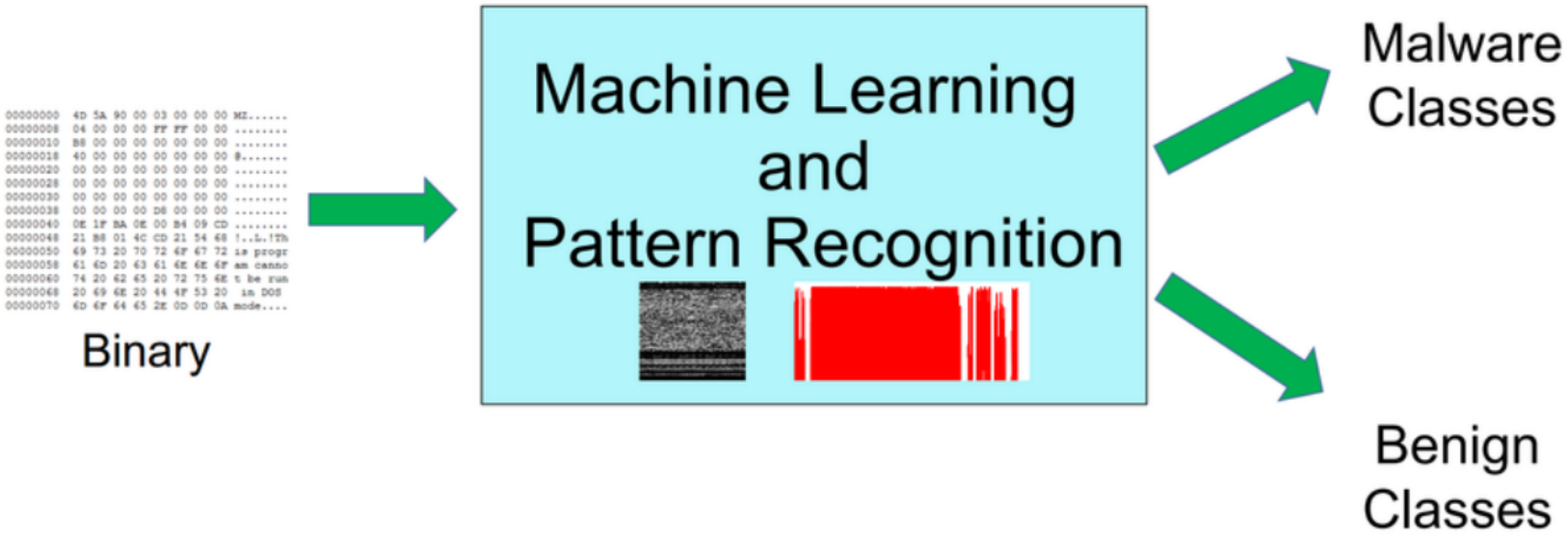

**Reference:** [Han+19]

# Traditional Signature-Based Malware Detection



# Malware Codes




Binary
Machine Learning
and
Pattern Recognition
Malware
Classes
Benign
Classes

# Table of Contents I

# Pseudo-code for image conversion procedure.

**Input:** Malware Image dataset $D_i | i = 1, 2....m$

**for** *each* $D_i | i = 1, 2....m$ **do**

Read file related to $D_i$

Calculate the size of the file.

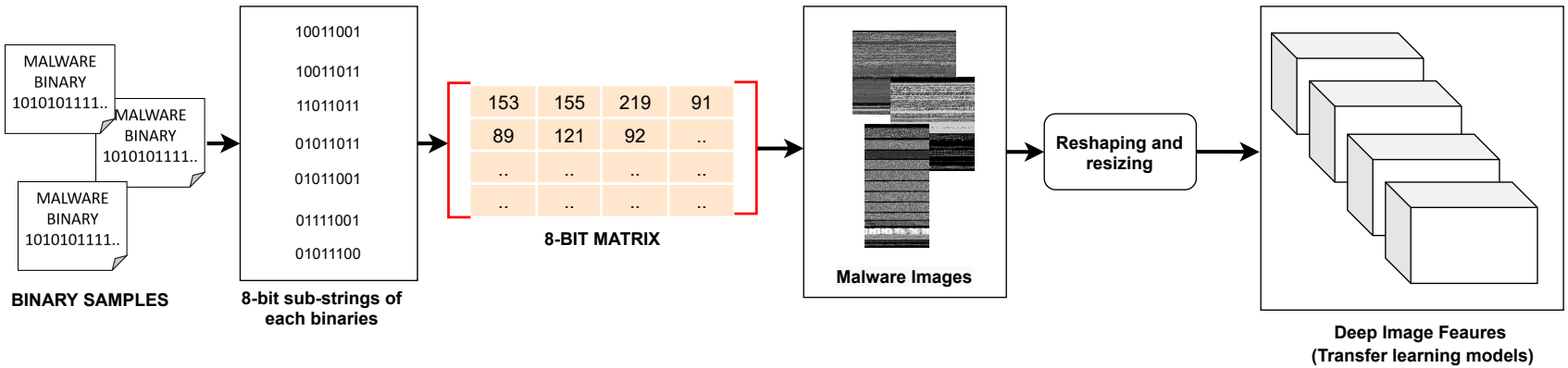
Initialize array,

Convert file into array of unsigned integer's $a$, between 0 and 255,

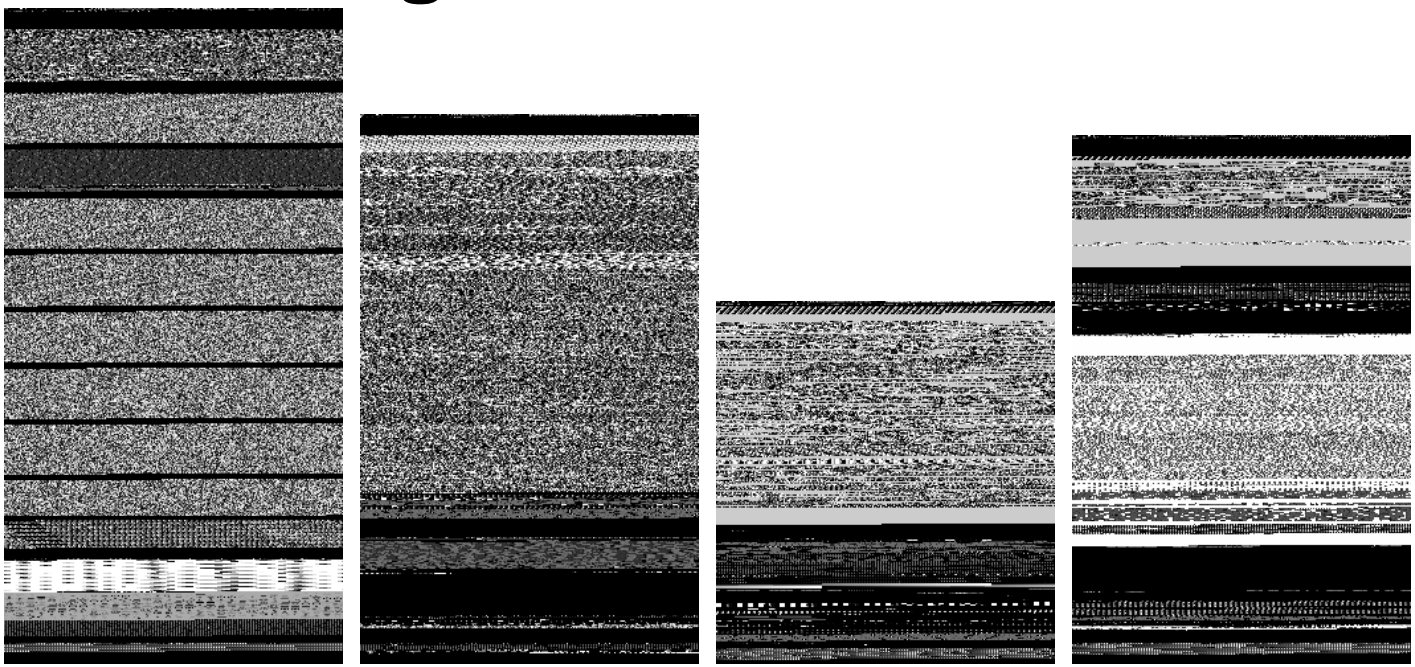
Convert array $a$ into $g$ as a 2-d array.

Convert the array $g$ into a grayscale image,

**end**

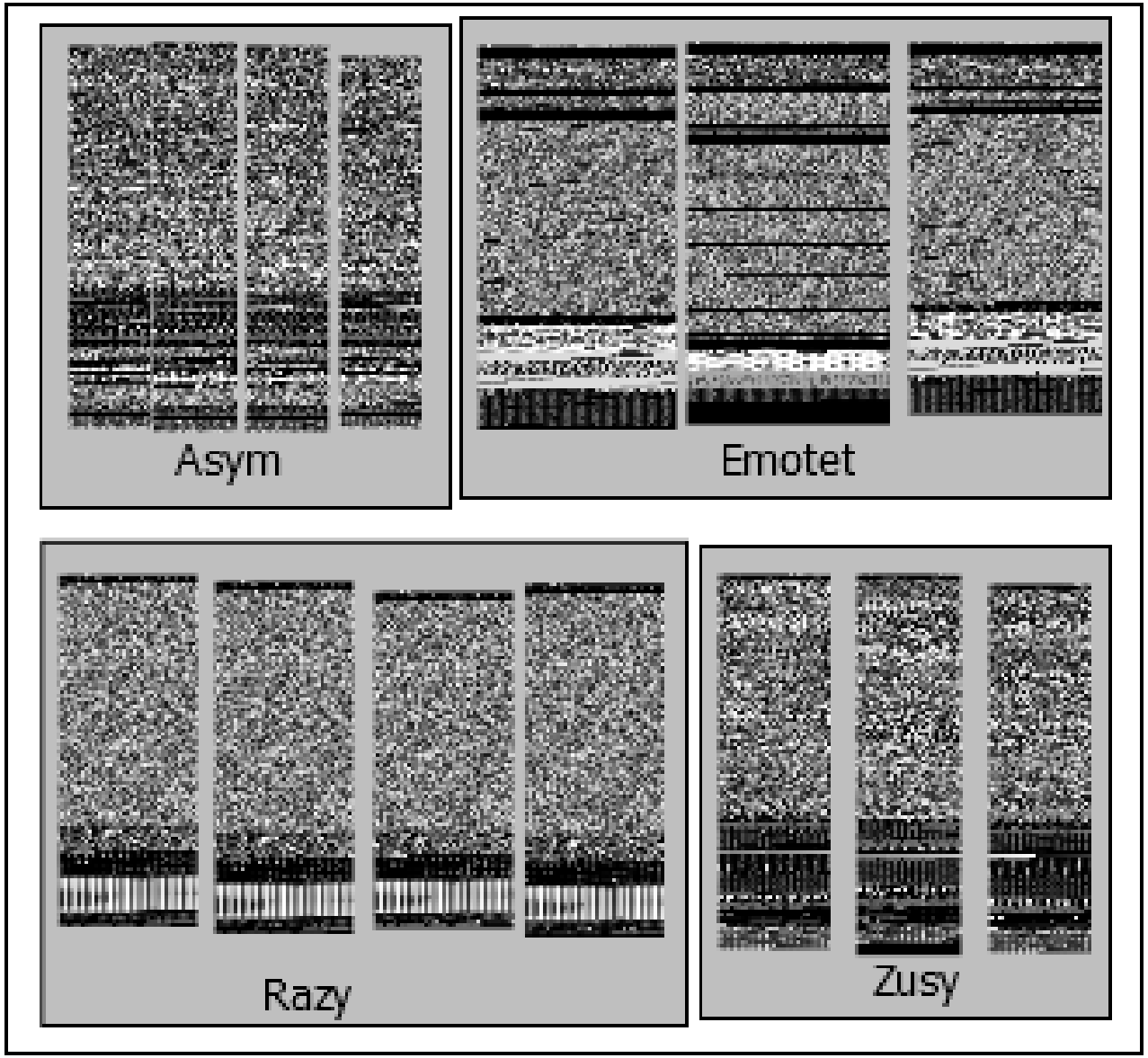


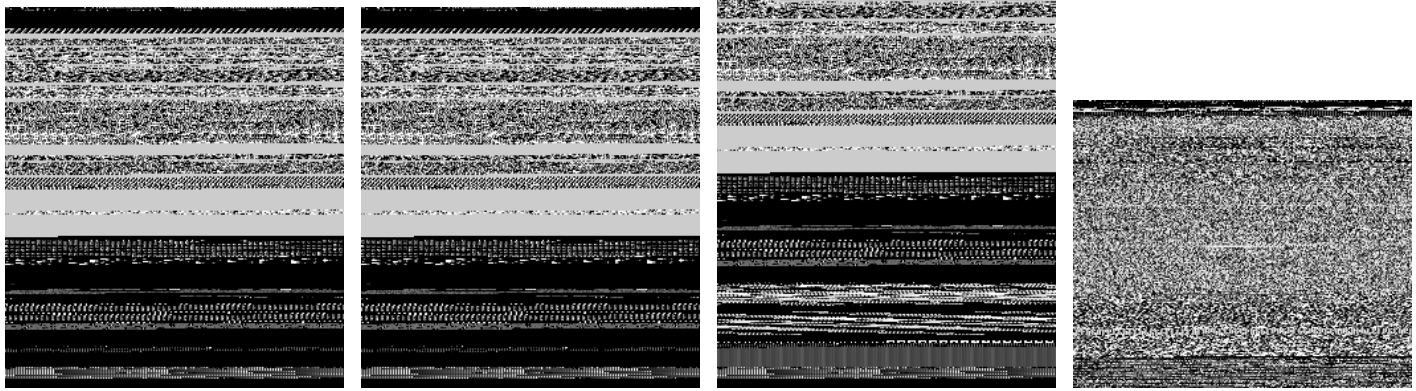
# Visualized images of a few real-world malware



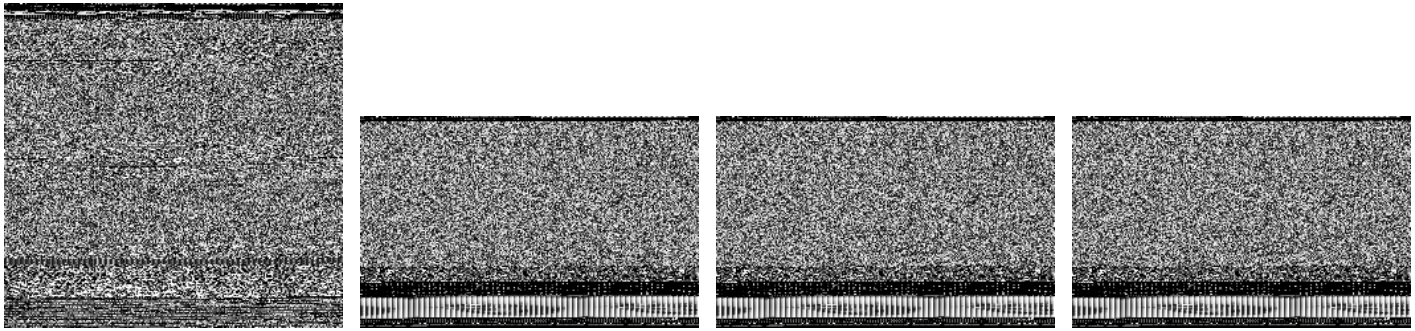
(a) Emotet. (b) Nitol. (c) Zusy. (d) Siscos.

# Similarity Analysis of same malware family



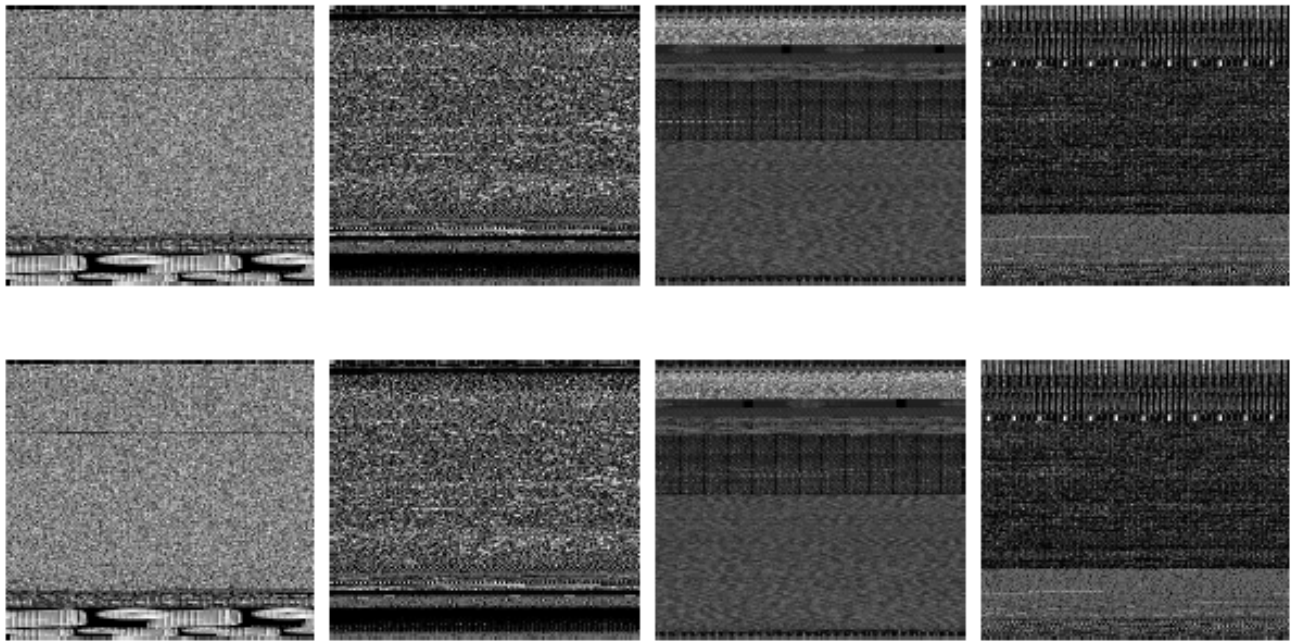


(a) UPX-farfli (b) UPX-farfli (c) UPX-farfli3 (d) UPX-nitol



(e) UPX-nitol (f) UPX-razy (g) UPX-razy (h) UPX-razy

# Examples of malware variants from different OS.



(a) Win (b) OS X (c) Linux (d) Android

Thank You!

**Contact: sanjeev@cdac.in, Mob: 9888751254**

# References I

[AAA20] Talal Abdullah, Waleed Ali, and Rawad Abdulghafor. "Empirical Study on Intelligent Android Malware Detection based on Supervised Machine Learning". In: *International Journal of Advanced Computer Science and Applications* 11 (2020).

[Afo+15] Vitor Monte Afonso et al. "Identifying Android malware using dynamically obtained features". In: *Journal of Computer Virology and Hacking Techniques* 11.1 (2015), pp. 9–17.

[Amī+20] Muḥammad Amīn et al. "Static malware detection and attribution in android byte-code through an end-to-end deep system". In: *Future Gener. Comput. Syst.* 102 (2020), pp. 112–126.

# References II

[AS15] M. Arefkhani and M. Soryani. "Malware clustering using image processing hashes". In: *2015 9th Iranian Conference on Machine Vision and Image Processing (MVIP)* (2015), pp. 214–218.

[Avd+15] Vitalii Avdiienko et al. "Mining apps for abnormal usage of sensitive data". In: *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*. Vol. 1. IEEE. 2015, pp. 426–436.

[Bay+09] Ulrich Bayer et al. "Scalable, Behavior-Based Malware Clustering". In: *NDSS*. Vol. 9. Citeseer. 2009, pp. 8–11.

[Bee+19a] C Beek et al. *McAfee Labs Threats Report August 2019*. https://www.mcafee.com/enterprise/en-us/assets/reports/rp-quarterly-threats-aug-2019.pdf. Accessed on 2020-24-09. Aug. 2019.

[Bee+19b] C Beek et al. "Mcafee labs threats report: August 2019". In: *McAfee Labs* (2019).

# References III

[Bho+19] Niket Bhodia et al. "Transfer Learning for Image-Based Malware Classification". In: *ArXiv* abs/1903.11551 (2019).

[Cai+18] Haipeng Cai et al. "DroidCat: Effective Android Malware Detection and Categorization via App-Level Profiling". In: *IEEE Transactions on Information Forensics and Security* 14.6 (2018), pp. 1455–1470.

[Cai20] Haipeng Cai. "Assessing and Improving Malware Detection Sustainability through App Evolution Studies". In: *ACM Transactions on Software Engineering and Methodology (TOSEM)* 29 (2020), pp. 1–28.

[Cui+18] Zhihua Cui et al. "Detection of Malicious Code Variants Based on Deep Learning". In: *IEEE Transactions on Industrial Informatics* 14 (2018), pp. 3187–3196.

[Cui+19] Zhihua Cui et al. "Malicious code detection based on CNNs and multi-objective algorithm". In: *J. Parallel Distributed Comput.* 129 (2019), pp. 50–58.

# References IV

[Das+16] Santanu Kumar Dash et al. "DroidScribe: Classifying Android Malware Based on Runtime Behavior". In: *2016 IEEE Security and Privacy Workshops (SPW)*. IEEE. 2016, pp. 252–261.

[DNB19] Venkata Salini Priyamvada Davuluru, Barath Narayanan Narayanan, and Eric J Balster. "Convolutional Neural Networks as Classification Tools and Feature Extractors for Distinguishing Malware Programs". In: *2019 IEEE National Aerospace and Electronics Conference (NAECON)*. IEEE. 2019, pp. 273–278.

[FC19] Xiaoqin Fu and Haipeng Cai. "On the Deterioration of Learning-Based Malware Detectors for Android". In: *2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)* (2019), pp. 272–273.

# References V

[Fer+20] Mohamed Amine Ferrag et al. "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study". In: *Journal of Information Security and Applications* 50 (2020), p. 102419.

[GRH10] John R Goodall, Hassan Radwan, and Lenny Halseth. "Visual analysis of code security". In: *Proceedings of the seventh international symposium on visualization for cyber security*. 2010, pp. 46–51.

[Han+19] Weijie Han et al. "MalInsight: A systematic profiling based malware detection framework". In: *Journal of Network and Computer Applications* 125 (2019), pp. 236–250.

[Jor+17] Roberto Jordaney et al. "Transcend: Detecting Concept Drift in Malware Classification Models". In: *26th {USENIX} Security Symposium ({USENIX} Security 17)*. 2017, pp. 625–642.

# References VI

[Kal+18] Mahmoud Kalash et al. "Malware Classification with Deep Convolutional Neural Networks". In: *2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)* (2018), pp. 1–5.

[Keb+17] Temesguen Messay Kebede et al. "Classification of malware programs using autoencoders based deep learning architecture and its application to the microsoft malware classification challenge (big 2015) dataset". In: *2017 IEEE National Aerospace and Electronics Conference (NAECON)*. IEEE. 2017, pp. 70–75.

[KJE19] Sanjeev Kumar, B Janet, and R Eswari. "Multi Platform Honeypot for Generation of Cyber Threat Intelligence". In: *2019 IEEE 9th International Conference on Advanced Computing (IACC)* (2019), pp. 25–29.

# References VII

[KJE20] Sanjeev Kumar, B Janet, and R Eswari. "Automated Cyber Threat Intelligence Generation from Honeypot Data". In: *Inventive Communication and Computational Technologies*. Springer, 2020, pp. 591–598.

[KM06] J Zico Kolter and Marcus A Maloof. "Learning to detect and classify malicious executables in the wild.". In: *Journal of Machine Learning Research* 7.12 (2006).

[KM13] Kesav Kancherla and Srinivas Mukkamala. "Image visualization based malware detection". In: *2013 IEEE Symposium on Computational Intelligence in Cyber Security (CICS)*. IEEE. 2013, pp. 40–44.

[KM22] Kambey L Kisambu and Mohamedi Mjahidi. "Evaluation of Machines Learning Algorithms in Detection of Malware-based Phishing Attacks for Securing E-Mail Communication". In: *CS & IT Conference Proceedings*. Vol. 12. 12. CS & IT Conference Proceedings. 2022.

# References VIII

[Kum+12] Sanjeev Kumar et al. "Distributed Honeynet System Using Gen III Virtual Honeynet". In: *International Journal of Computer Theory and Engineering* 4.4 (2012), p. 537.

[LNP15] Martina Lindorfer, Matthias Neugschwandtner, and Christian Platzer. "MARVIN: Efficient and comprehensive mobile app classification through static and dynamic analysis". In: *2015 IEEE 39th annual computer software and applications conference*. Vol. 2. IEEE. 2015, pp. 422–433.

[Nae+20] Hamad Naeem et al. "Malware detection in industrial internet of things based on hybrid image visualization and deep learning model". In: *Ad Hoc Networks* 105 (2020), p. 102154.

# References IX

[Nat+11a] Lakshmanan Nataraj et al. "A comparative assessment of malware classification using binary texture analysis and dynamic analysis". In: *Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence*. 2011, pp. 21–30.

[Nat+11b] Lakshmanan Nataraj et al. "Malware images: visualization and automatic classification". In: *Proceedings of the 8th international symposium on visualization for cyber security*. 2011, pp. 1–7.

[ND20] Barath Narayanan Narayanan and Venkata Salini Priyamvada Davuluru. "Ensemble Malware Classification System Using Deep Neural Networks". In: *Electronics* 9.5 (2020), p. 721.

# References X

[NDK16] Barath Narayanan Narayanan, Ouboti Djaneye-Boundjou, and Temesguen M Kebede. "Performance analysis of machine learning and pattern recognition algorithms for malware classification". In: *2016 IEEE National Aerospace and Electronics Conference (NAECON) and Ohio Innovation Summit (OIS)*. IEEE. 2016, pp. 338–342.

[NQZ18] Sang Ni, Quan Qian, and Rui Zhang. "Malware identification using visualization images and deep learning". In: *Computers & Security* 77 (2018), pp. 871–885.

[RFT16] Paulo E. Rauber, A. Falcão, and A. Telea. "Visualizing Time-Dependent Data Using Dynamic t-SNE". In: *EuroVis*. 2016.

[Rie+11] Konrad Rieck et al. "Automatic analysis of malware behavior using machine learning". In: *Journal of Computer Security* 19.4 (2011), pp. 639–668.

# References XI

[Ron+18] R. Ronen et al. "Microsoft Malware Classification Challenge". In: *ArXiv* abs/1802.10135 (2018).

[SB15] Joshua Saxe and Konstantin Berlin. "Deep neural network based malware detection using two dimensional binary program features". In: *2015 10th International Conference on Malicious and Unwanted Software (MALWARE)*. IEEE. 2015, pp. 11–20.

[SC20] Silvia Sebasti'an and Juan Caballero. "AVclass2: Massive Malware Tag Extraction from AV Labels". In: *Annual Computer Security Applications Conference* (2020).

[Sch+01] Matthew G Schultz et al. "Data mining methods for detection of new malicious executables". In: *Proceedings 2001 IEEE Symposium on Security and Privacy. S&P 2001*. IEEE. 2001, pp. 38–49. DOI: 10.1109/SECPRI.2001.924286.

# References XII

[Sha+11] Asaf Shabtai et al. "Detecting unknown malicious code by applying classification techniques on OpCode patterns". In: *Security Informatics* 1 (2011), pp. 1–22.

[SS15] PV Shijo and AJPCS Salim. "Integrated static and dynamic analysis for malware detection". In: *Procedia Computer Science* 46 (2015), pp. 804–811.

[Sua+17] Guillermo Suarez-Tangil et al. "Droidsieve: Fast and accurate classification of obfuscated android malware". In: *Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy*. 2017, pp. 309–320.

[Sym20] Symantec. *symantec threat landscape report 2020*. https://symantec-enterprise-blogs.security.com/blogs/threat-intelligence/threat-landscape-q1-2020. Accessed on 2020-24-09. June 2020.

# References XIII

[Tam+15] Kimberly Tam et al. "CopperDroid: Automatic Reconstruction of Android Malware Behaviors". In: *Ndss*. 2015.

[Tri+09] Philipp Trinius et al. "Visual analysis of malware behavior using treemaps and thread graphs". In: *2009 6th International Workshop on Visualization for Cyber Security* (2009), pp. 33–38.

[Vas+20a] D. Vasan et al. "Image-Based malware classification using ensemble of CNN architectures (IMCEC)". In: *Comput. Secur.* 92 (2020), p. 101748.

[Vas+20b] D. Vasan et al. "IMCFN: Image-based malware classification using fine-tuned convolutional neural network architecture". In: *Comput. Networks* 171 (2020), p. 107138.

# References XIV

[VAV19] Sitalakshmi Venkatraman, Mamoun Alazab, and R Vinayakumar. "A hybrid deep learning image-based analysis for effective malware detection". In: *Journal of Information Security and Applications* 47 (2019), pp. 377–389.

[Vin+19] R. Vinayakumar et al. "Robust Intelligent Malware Detection Using Deep Learning". In: *IEEE Access* 7 (2019), pp. 46717–46738.

[Wag+15] Markus Wagner et al. "A Survey of Visualization Systems for Malware Analysis". In: *Eurographics Conference on Visualization (EuroVis) - STARs*. Ed. by R. Borgo, F. Ganovelli, and I. Viola. The Eurographics Association, 2015. DOI: 10.2312/eurovisstar.20151114.

# References XV

[Xia+20] Guoqing Xiao et al. "MalFCS: An effective malware classification framework with automated feature extraction based on deep convolutional neural networks". In: *J. Parallel Distributed Comput.* 141 (2020), pp. 49–58.

[Xu+19] K. Xu et al. "DroidEvolver: Self-Evolving Android Malware Detection System". In: *2019 IEEE European Symposium on Security and Privacy (EuroS&P)* (2019), pp. 47–62.

[Yoo04] InSeon Yoo. "Visualizing windows executable viruses using self-organizing maps". In: *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*. 2004, pp. 82–89.

[Zha+16] Jixin Zhang et al. "IRMD: Malware Variant Detection Using Opcode Image Recognition". In: *2016 IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS)* (2016), pp. 1175–1180.

# References XVI

[Zha+20] Xiaohan Zhang et al. "Enhancing State-of-the-art Classifiers with API Semantics to Detect Evolved Android Malware". In: *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security* (2020).

[ZJ11] Mohamad Fadli Zolkipli and Aman Jantan. "An approach for malware behavior identification and classification". In: *2011 3rd International Conference on Computer Research and Development*. Vol. 1. IEEE. 2011, pp. 191–194.