A Survey on Ransomware

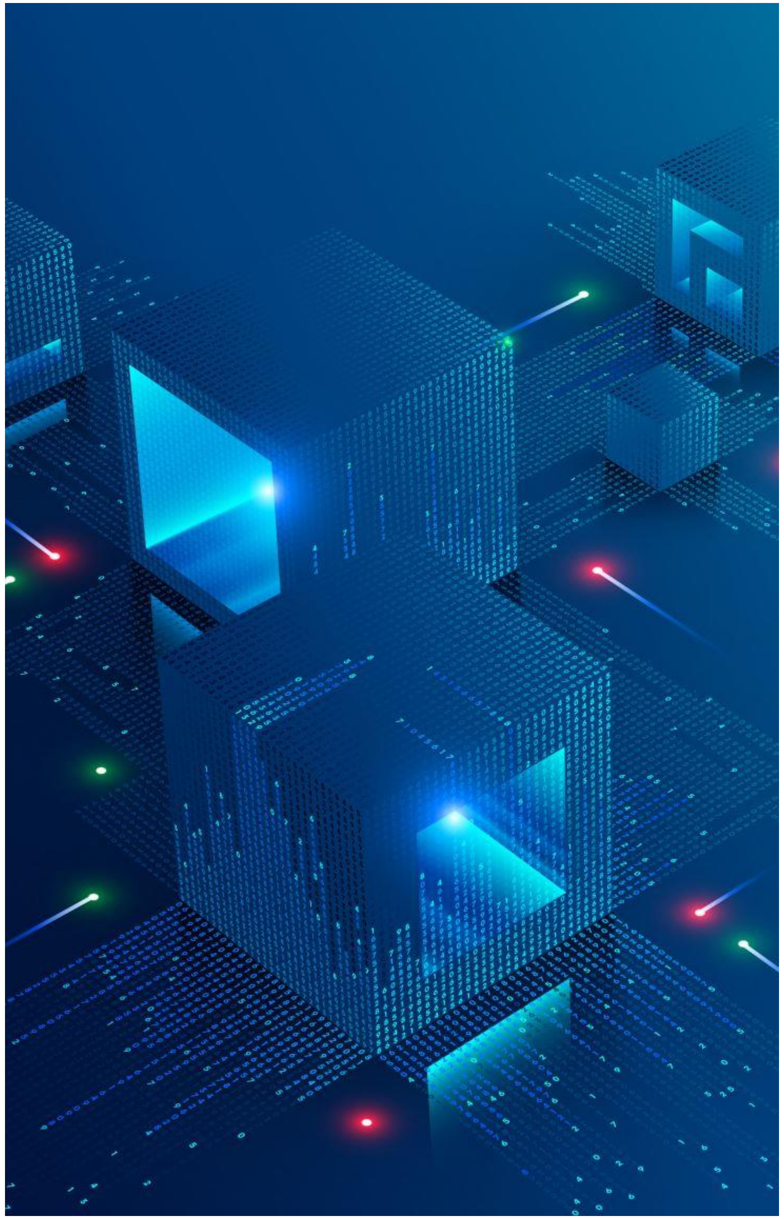
Evolution, Taxonomy, and Defense Solutions

Written By: Harun Oz and Ahmet Aris
(Florida International University, USA)

Presented By: 제레미아

Course: Advanced Security in Emerging ICT

Monday, September 19, 2022



PART 2 – DEFENSE, OPEN ISSUES & CONCLUSION

5. Ransomware Defense Research

   5.1 Ransomware Analysis Research

   5.2 Ransomware Detection Research

   5.3 Ransomware Recovery Research

   5.4 Other Ransomware Defense Research

6. Open Issues
   - ❑ Constant Evolution of Ransomware
   - ❑ Human operated ransomware attacks
   - ❑ Rootkit Fashion
   - ❑ Ransomware Living of the land
   - ❑ Changing encryption tradition
   - ❑ Leveraging internal attacks
   - ❑ New ransomware targets
   - ❑ Willingness to pay (And many more OI's)
7. Conclusion
8. References

RANSOMWARE DEFENSE OVERVIEW IN LITERATURE

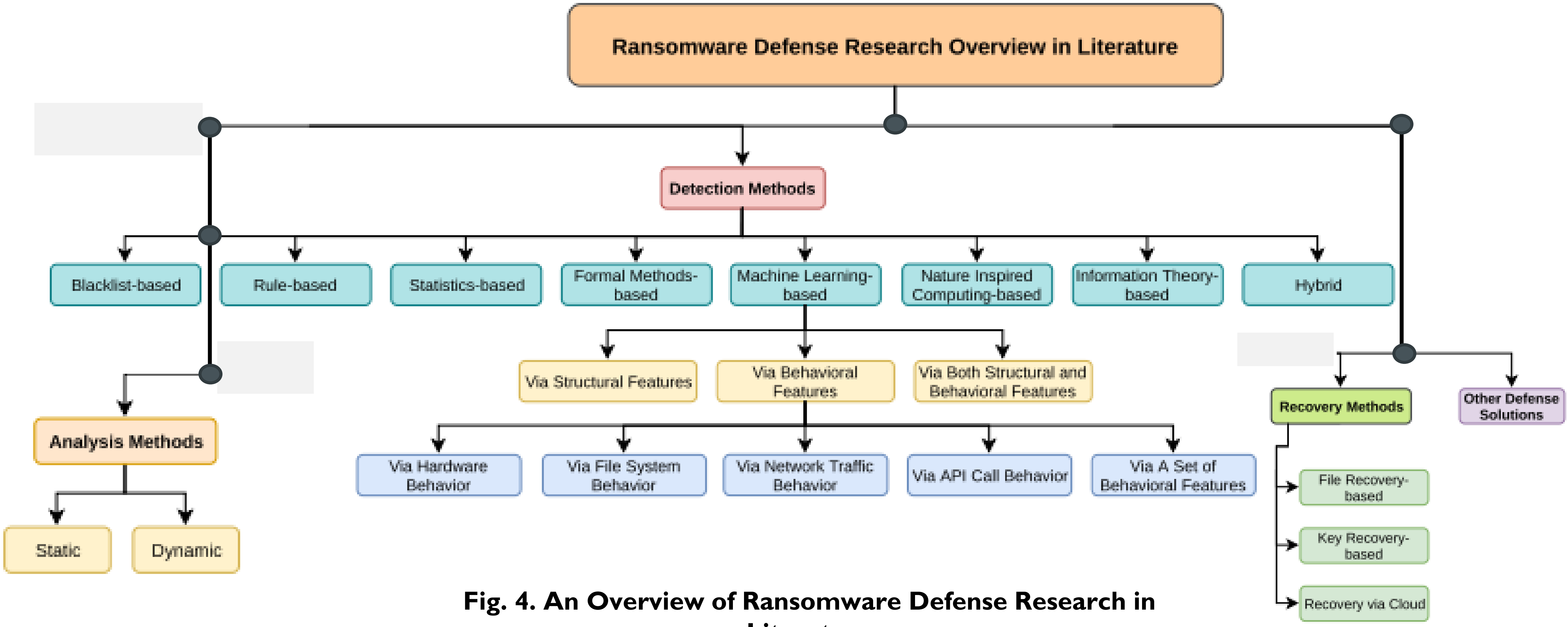


Fig. 4. An Overview of Ransomware Defense Research in Literature

5.

RANSOMWARE DEFENSE RESEARCH

Ransomware Defense Research Categories:

- ❑ Analysis, Detection, Recovery and Other defense research are the four main categories of ransomware defense.
- ❑ For this study, the authors provided a taxonomy of each research domain with respect to target platform, i.e.,
  - ✓ PCs/workstations,
  - ✓ Mobile Devices, and
  - ✓ IoT/CPS
- ❑ Based on the target platforms, authors did the following:
  - ✓ Give an overview of various ransomware analysis techniques
  - ✓ Categorize and explain ransomware detection systems
  - ✓ Finally summarize the recovery mechanisms
  - ✓ However, there exist some studies that do not fall into any of the aforementioned categories and therefore were summarized under Other Methods category in this survey

5.1 Ransomware Analysis Research:

- ❑ Ransomware analysis includes activities to understand the behavior and/or characteristics of ransomware. Similar to traditional malware analysis, ransomware analysis techniques can be categorized as **static** and **dynamic**.
- ❑ **Static analysis** aims to understand whether a sample is a ransomware or not by extracting structural information from the sample without actually running it.
- ❑ **Dynamic analysis** of ransomware consists of running the sample and observing the behavior to determine if the sample is a ransomware or not.
  - ✓ **Dynamic analysis** is performed via running the samples inside an isolated environment (i.e., sandbox) to avoid a possible damage caused by the analyzed sample
- ❑ Static and Dynamic analysis have their own advantages and disadvantages, which result in researchers to use both of the approaches resulting into **Hybrid Analysis**

5.1.1 Ransomware Analysis in PC/Workstations:

- ✓ **Structural** and **Behavioral features** obtained via static and dynamic analysis of ransomware samples targeting PCs/workstations, respectively
- ✓ **Structural features** obtained from ransomware for PCs/workstations consist of file hashes, header information, function/API/system calls, strings, opcodes, and file types.
- ✓ **Behavioral features** obtained from ransomware for PCs/workstations include registry activity, host logs, process activity, file system activity, inputs and outputs of function/API/system calls, I/O access patterns, network activity, resource usage, and sensor readings.

5.

RANSOMWARE DEFENSE RESEARCH

5.1.2 Ransomware Analysis in Mobile Devices:

- ❑ This subsection, gives an overview of structural and behavioral features obtained from static and dynamic analysis of ransomware samples targeting mobile devices, respectively.
- ❑ Structural features obtained from ransomware for mobile devices are strings, opcodes, application images, permissions requests and API packages.
  - ✓ Strings: The strings that are extracted from the packaged mobile application can be used as a feature to detect mobile ransomware. Such strings can contain IP addresses, domain names, ransom notes, etc., which can be helpful to detect ransomware.
  - ✓ Opcodes: Instruction opcodes that are obtained from the disassembled application byte-code can be used to understand if a mobile application has the characteristics of ransomware
  - ✓ Application Images: Extracted images from the application may contain ransom related material (i.e., ransom message image), and thus be used as a feature to detect mobile ransomware.
  - ✓ Permissions: Mobile applications require permissions to be approved by the users to access and utilize resources of the mobile device.
  - ✓ API Packages: API packages can be extracted from the source code of a mobile application to determine the malicious encryption or locking characteristics.

5.1.2 Ransomware Analysis in Mobile Devices:

- ❑ Behavioral features obtained from ransomware for mobile devices are function/API/system calls, user interaction, file system features, and resource usage.
  - ✓ Function/API/system calls: Researchers can detect mobile ransomware variants by analyzing the function/API/system calls made by a mobile application while running.
  - ✓ User Interaction: Matching the user's interactions with the events taking place while the application is running can be used to detect the presence of a ransomware.
  - ✓ File System Features: Like in PCs/workstations, the features extracted from the file system of a mobile device can be used to understand the presence of ransomware.
  - ✓ Resource Usage: Similar to PCs/workstations, abnormalities in the resource usage patterns on amobile device, such as power consumption, can be a sign of the presence of a mobile ransomware.

5.1.3 Ransomware Analysis in IoT/CPS Platforms:

- ❑ This section, gives an overview of structural and behavioral features extracted from ransomware that can target IoT/CPS platforms. However, defense research for IoT/CPS environments is in its **infancy** at the moment so only few studies exist in the literature, and therefore, only behavioral features, namely, **network activities** were used in the literature.

  - ✓ **Network Activity:** Network-related features are captured by researchers within the IoT/CPS environment to find out the communication patterns signifying the presence of ransomware.

5.2

RANSOMWARE DETECTION RESEARCH

5.1 Ransomware Detection Research

- In this subsection, authors categorize and summarize existing **detection mechanisms** for ransomware with respect to **target platforms**. Based on the employed methodology, they categorize detection systems into eight categories:

1. **Blacklist-based**: the system detects ransomware using a list of malicious domain names or IP addresses that are known to be used by ransomware families.
2. **Rule-based:** the system detects ransomware using rules that are constructed using the analysis features. Rules can be either the rules compatible with malware detection engines (e.g., YARA), maliciousness scores, or threshold values.
3. **Statistics-based:** the system detects ransomware using statistics on features indicating that the sample is a ransomware.
4. Formal Methods-based: the system detects ransomware using a formal model that can discriminate malicious and benign patterns.
5. **Nature Inspired Computing-based**: the system detects ransomware using techniques inspired from the nature and biology.
6. **Information Theory-based**: the system detects ransomware using information theory approaches (e.g., entropy). Encryption operation performed by cryptographic ransomware strains results in changes in the information content of the files.
7. **Machine Learning-based:** the system detects ransomware via ML models that are built using a set of analysis features. ML-based ransomware detection systems use either structural features, behavioral features, or both. Structural features are obtained by researchers via static analysis of ransomware binaries.
8. **Hybrid:** the system detects ransomware via a set of the detection techniques

5.2.1 Ransomware Detection for PCs/Workstations:

1. **Blacklist-Based Detection**. Akbanov et al. [10] examined the behavior of WannaCry ransomware on SDN and proposed an SDN-based ransomware detection method. Their detection system runs as an application on the SDN controller and monitors the network traffic for the appearance of malicious domain names or the IP addresses used by WannaCry. Once a matching flow is detected, rules to block that malicious traffic are generated.

2. **Rule-Based Detection**. YARA rules are created by the rule-based ransomware detection system of Medhat et al. [126] using API calls of file and cryptography libraries, strings, and file extensions from ransomware binaries.

3. **Statistics-Based Detection**. Palisse et al. proposed a statistics-based ransomware detection system, namely Data Aware Defense (DAD) [141]. DAD focuses on features obtained from write operations such as buffer content, size, offset, file name, process id and name, and thread id.

4. **Information Theory-Based Detection**. Since benign encryption, compression, and file conversion operations on already compressed file formats also result in high entropy values, several researchers used entropy as a supportive feature for their detection systems.

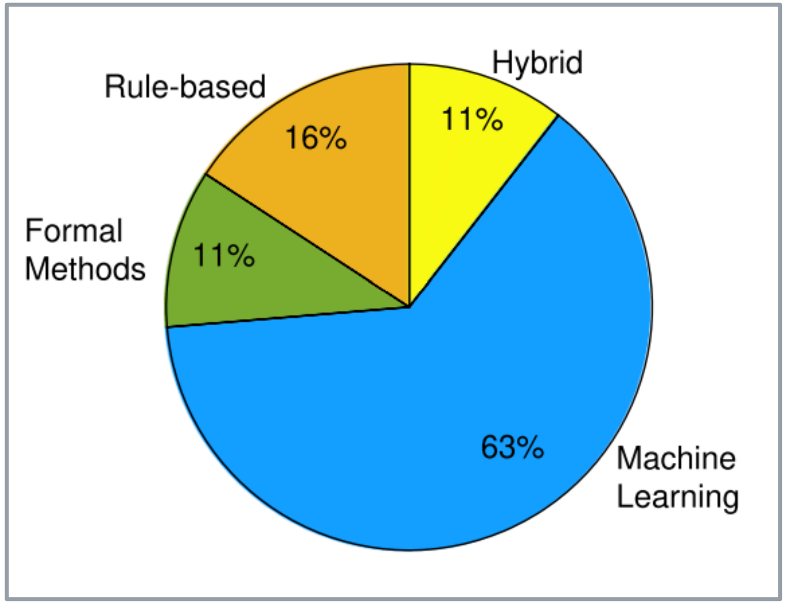
5.2.1 Ransomware Detection for PCs/Workstations:

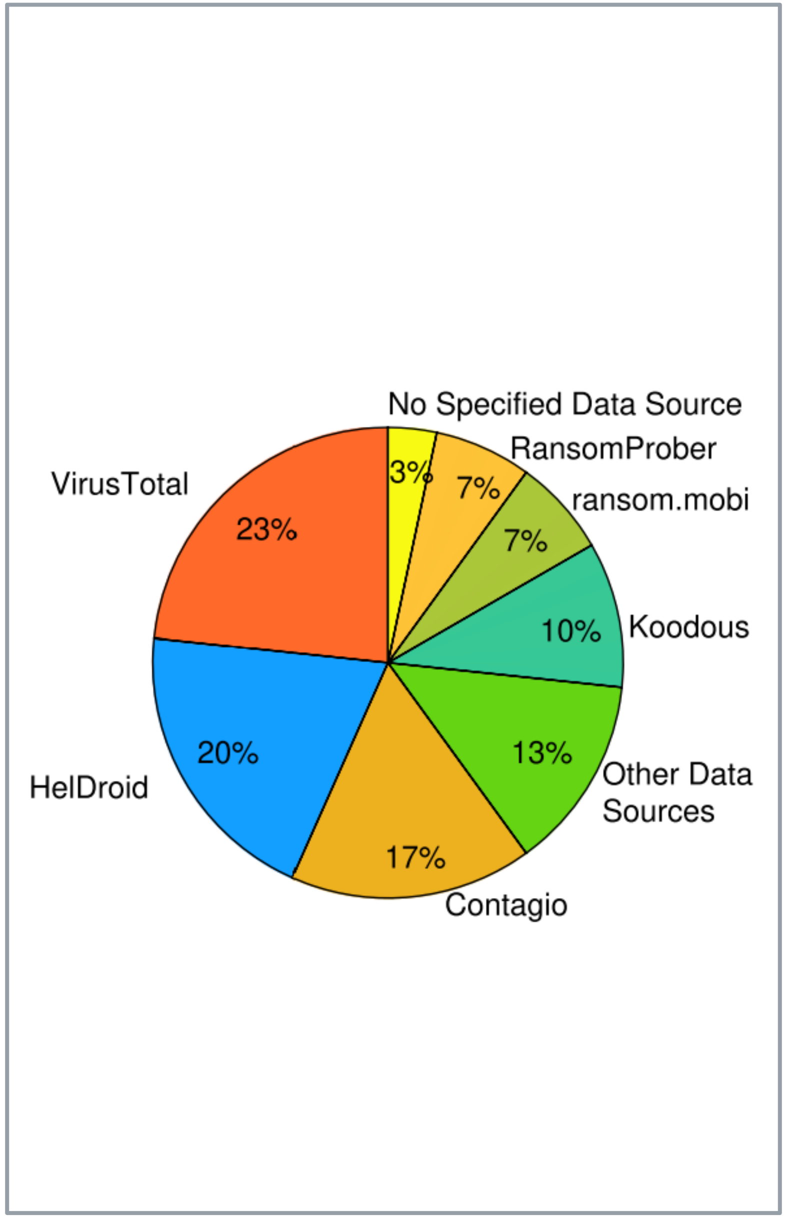
5. **Machine Learning-Based Detection**.

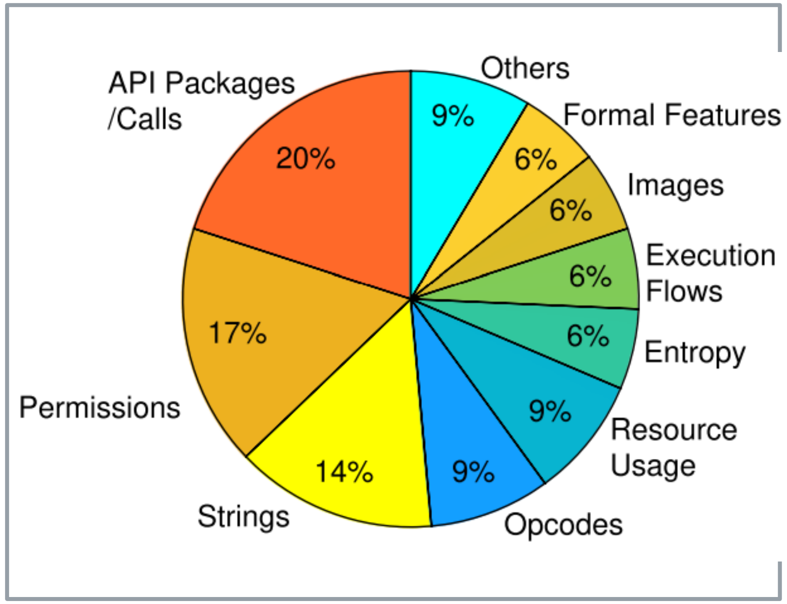
- ✓ **Via Structural Features:** In terms of the ML-based ransomware detection systems for PCs/workstations using structural features, researchers employed instruction opcodes, API calls, and DLLs.
- ✓ **Via Behavioral Features:** In terms of the ML-based ransomware detection systems proposed for PCs/workstations using behavioral features, researchers monitored and/or analyzed hardware, file system, network traffic, and API call behaviors.
- ✓ **Hardware Behavior:** PC/workstation hardware including storage hardware, on-board sensors, and memory dumps were monitored by researchers for ransomware detection.
- ✓ **Via File System Behavior:** Instead of monitoring the hardware, some researchers aimed to detect ransomware at a higher level via monitoring file system activities. Compared to hardware behavior, file system behavior monitoring can provide a lower granular data allowing to obtain file and process information.
- ✓ **Via Network Traffic Behavior:** Since ransomware usually communicates with its C&C server for key exchange or data exfiltration, some researchers aimed to detect ransomware in the networked devices by observing the network traffic.
- ✓ **Via API Call Behavior**: One of the main behavioral features obtained from dynamic analysis of ransomware is API calls. In this context, some studies used API calls as features to build ML classifiers to detect ransomware in PCs/workstations.
- ✓ **Via a Set of Behavioral Features:** Some of the studies used a set of behavioral features to build ML classifiers to detect ransomware in PCs/workstations. In this regard, a Bayesian Belief Network (BBN) classifier, an LSTM classifier, and multiple ML classifiers were built for ransomware detection.
- ✓ **Via Both Structural and Behavioral Features:** Instead of using only structural or behavioral features, some of the researchers employed features from both groups for ransomware detection. Artificial Neural Networks (ANNs) and SVM classifiers,

5.2.1 Ransomware Detection for PCs/Workstations:

6. **Formal Methods-Based Detection**. In [91], Iffländer et al. proposed DIMAQS (Dynamic Identification of Malicious Query Sequences) for detection of ransomware targeting database servers. DIMAQS utilizes colored Petri nets-based classifier to detect the malicious query sequences made by ransomware to target database servers.
7. **Nature Inspired Computing-Based Detection**. An Artificial Immune System-based ransomware detection system was proposed by Lu et al. [116]. The proposed system uses API call n-grams as antigens and employs a double-layer negative selection algorithm to discriminate ransomware from benign applications.
8. **Hybrid Detection**. In addition to the studies employing one of the aforementioned detection techniques, a few studies exist in the literature that used a set of those approaches.







OVERVIEW OF RANSOMWARE DETECTION RESEARCH FOR PCS/WORKSTATIONS

Overview:

- ✓ **Detection Techniques:** ML-Based detection is the most widely used approach for ransomware detection for PCs/workstations (73%), majority of the studies used behavioral features (43%), structural features (12%) and both features (12%)

- ✓ **Detection Features:** API calls and file/directory features are the most popular features used for ransomware detection for PCs/workstations.

- ✓ **Evaluation Datasets:** VirusTotal is the most popular data source for ransomware detection systems for PCs/workstations. It is followed by VirusShare, hybridanalysis.com, and the others.

5.2.2 Ransomware Detection for Mobile Devices

❑ This subsection categorizes and give an overview of ransomware detection systems for mobile devices. Considering the existing works, rule-based, formal methods-based, machine learning-based, and hybrid detection techniques were employed by researchers

1. **Rule-Based Detection**. Three rule-based mobile ransomware detection systems were proposed by researchers that use threshold values for detection. RanDroid [24] extracts images and strings from applications and calculates their similarity to the images and strings of ransomware samples. Based on the threshold values, it detects mobile ransomware
2. **Formal Method-Based Detection**. Formal methods to detect mobile ransomware were employed by two studies in the literature. The defense solution proposed in [129] and its extended version in [50] leveraged Calculus of Communicating Systems (CCS) formal model to detect mobile ransomware.
3. **Machine Learning-Based Detection** - (Discussed in detail in the next slide)
4. **Hybrid Detection**. In addition to the studies employing only one of the aforementioned detection techniques, a few studies exist in the literature that used a set of those approaches.

5.2.2 Ransomware Detection for Mobile Devices

3. Machine Learning-Based Detection.

- ✓ **Via Structural Features:** In terms of the ML-based ransomware detection systems for mobile devices using structural features, researchers used API packages [20, 121], classes, and methods [157], permissions [21], opcodes in native instruction formats [111], grey-scale images of mobile application source codes [98], and structural entropy of mobile applications [57] to build and evaluate various ML classifiers. Some researchers aimed to offload the mobile ransomware detection tasks to cloud to save from the resources of mobile devices.
- ✓ **Hardware Behavior:** Power usage behavior of mobile applications was used by Azmoodeh et al. [32] to detect ransomware. They used PowerTutor application to collect power consumption of both benign and ransomware applications at regular intervals, and analyzed the performance of a number of ML classifiers on the collected data
- ✓ Via Both Structural and Behavioral Features: A few studies in the literature aimed to benefit from both static and dynamic analysis of mobile ransomware samples and use the obtained features to build ML models. Ferrante et al. [67] proposed a mobile ransomware detection system that extracts opcode frequencies via static analysis and obtains CPU, memory, network usage, and system call statistics via dynamic analysis. In total, 87 features were used to train and evaluate various ML classifiers.

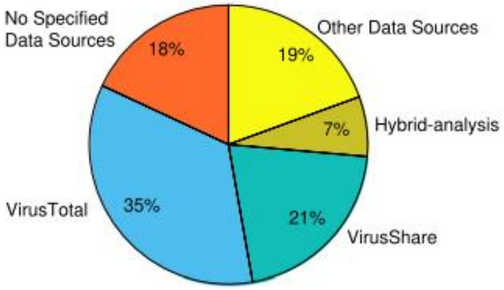
OVERVIEW OF RANSOMWARE DETECTION RESEARCH FOR MOBILE DEVICES

Overview:

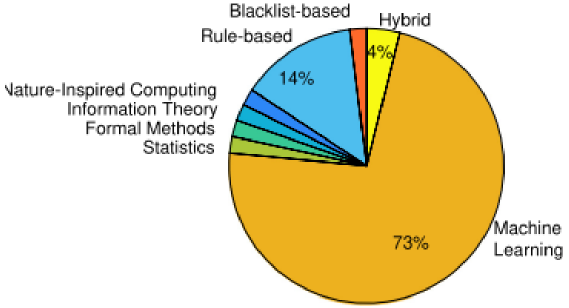
- ✓ **Detection Techniques & Features:** ML-Based detection is the most widely used approach for ransomware detection in mobile devices (over 60%), majority of the studies used structural features obtained visa static analysis to build ML Models.

- ✓ **Evaluation Datasets:** The most popular data source for ransomware detection systems for mobile devices are VirusTotal and the dataset from HelDroid.

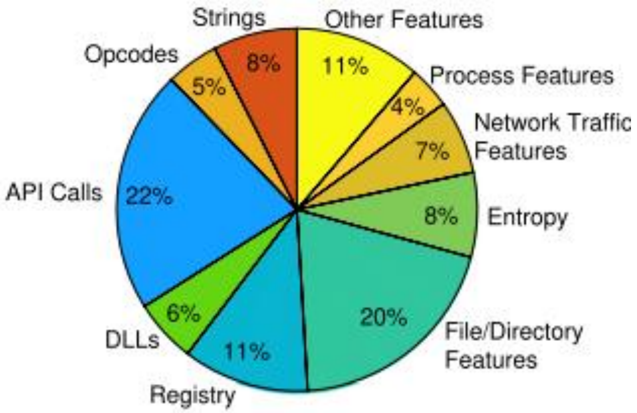
- ✓ **Detection Accuracy:** The ransomware detection studies for mobile devices reported very high detection rates. TPR changes between 83% and 100%, while FPR varies between 0 and 19%. Only one study reported a perfect TPR (i.e., 100%), while several studies reported a TPR over 99%.



(c) Evaluation Dataset



(a) Detection Techniques



(b) Detection Features

5.2.3 Ransomware Detection for IoT/CPS

- ❑ Since ransomware detection for IoT/CPS environments is not a well explored field of research, there are only five studies tackling the ransomware detection problem in such environments. Considering the detection studies, all of the studies utilize ML techniques.

1. **Machine Learning-Based Detection**.

   - ✓ **Via Network Traffic Behavior**: Considering the ML-based ransomware detection systems for IoT/CPS, there exist two studies. In the first study, Maimó et al. [66] proposed a ransomware defense system for Integrated Clinical Environments (ICE) of Medical CPS.

   - ✓ **Via a Set of Behavioral Features:** Al-Hawawreh and Sitnikova [14] proposed a DL-based ransomware detection system for the workstations that are used as host machines of Industrial IoT environments. Their system relies on classical and variational auto-encoders to select the most appropriate features from several behavioral features of API calls, registry keys, file and directory operations.

5.2.4 Comparison of Ransomware Detection Techniques Across All Platforms

- This subsection, compares the detection studies in PCs/workstations, mobile devices, and IoT/CPS environments and share findings with ransomware detection across various platforms.
  - ✓ **Comparison of the Detection Techniques:** Analysis disclosed that machine learning is the most admired technique to detect ransomware across all platforms. Specifically, in total 72% of defense solutions utilized machine learning to detect ransomware in the system.
  - ✓ **Comparison of the Used Features:** Findings show that ransomware detection studies for PCs/workstations and IoT/CPS environments display a different behavior than the studies for mobile devices. Specifically, we see that majority of the machine learning-based ransomware detection systems for PCs/workstations and IoT/CPS environments rely on behavioral features. Whereas most of the studies for mobile devices utilize structural features. In general, structural features are easier are easier to extract/collect compared to behavioral features as they do not require samples to run and do not necessitate monitoring of the platform.
  - ✓ **Comparison of the Datasets:** The most widely used data source for ransomware detection systems across all platforms is VirusTotal. This finding is not surprising as VirusTotal is a very popular repository for malware research domain, and it provides an academic dataset and an API to researchers from academia free of charge.
  - ✓ **Comparison of the Detection Accuracy:** Generally, all of the reviewed ransomware detection studies reported very high detection rates. Specifically, while TPR fluctuates between 73% and 100%, FPR changes between 0 and 19%.

5.3 Ransomware Recovery Research:

- ❑ This section categorize and summarize existing recovery mechanisms for ransomware with respect to target platforms.

5.3.1 Ransomware Recovery for PC/Workstations:

- ❑ Ransomware recovery research for PCs/ workstations shows that recovery of the destruction performed by ransomware can be achieved in three different ways: recovery of keys, recovery of files via hardware, or recovery of files via cloud backup.

  - ✓ **Recovery of Keys:** a key-escrow mechanism has been proposed that intends to capture encryption key(s) by hooking the cryptography APIs and decrypt the victim files. Naturally, it is effective only against the ransomware families that call the corresponding cryptography APIs for encryption.

  - ✓ **Recovery of Files via Hardware:** The studies presented in this category aim to recover encrypted files of victims by utilizing the characteristics of storage hardware (i.e., SSD). NAND-based SSDs have the ability of out-of-place update feature that preserves a previous version of deleted data until the Garbage Collector (GC) deletes it.

  - ✓ **Recovery of Files via Cloud Backup:** Some of the recovery mechanisms in the literature aimed to recover files utilizing cloud environment for backup purposes. Yun et al. [199] proposed a backup system named CLDSafe that is deployed on the cloud.

5.3 Ransomware Recovery Research:

5.3.2 Ransomware Recovery for Mobile Devices:

- ❑ Considering the recovery solutions for mobile devices to enable data recovery from ransomware attacks, there exists:
  - ✓ MimosaFTL [189] was designed as a recovery-based ransomware defense strategy for mobile devices that are equipped with flash memory as external storage. It collects the access behaviors of ransomware samples and applies K-mean clustering to identify the unique access patterns to the Flash Transaction Layer.
  - ✓ **In [59] Yalew et al.** aimed to recover from ransomware by periodically performing backups to an external storage.a backup system named CLDSafe that is deployed on the cloud.

5.

RANSOMWARE DEFENSE RESEARCH

5.4 Other Ransomware Defense Research:

- ❑ Ransomware defense is a very active topic of research. In this subsection we give a brief overview of rest of the defense studies that do not fall under the categorization applied earlier. These studies can be grouped into moving target, access control, and holistic defense categories
  - ✓ **Moving target defense** technique was proposed by Lee et al. [114] for ransomware protection that changes the file extensions randomly.
  - ✓ In terms of the **access control mechanisms**, Genç et al. [75] proposed UShallNotPass that aims to prevent ransomware attack before performing encryption by blocking the access of unauthorized applications to the pseudo-random number generator functions in the operating system.
  - ✓ Considering the **holistic defense systems,** Keong et al. proposed VoterChoice [99] that uses Suricata Intrusion Prevention System to detect malicious activities. Once such an activity is detected, ML-based detection modules that use encryption and registry activities as features detect ransomware.

7. CONCLUSION

Concluding Remarks.

- This paper provided a comprehensive survey of ransomware and ransomware defense research with respect to PCs/workstations, mobile devices and IoT/CPS environments.
- Moreover, the study presented how a detailed overview on how ransomware evolved in time, thoroughly analyzed the key building blocks of ransomware, proposed a taxonomy of notable ransomware families, and provided an extensive overview of ransomware defense research including analysis, detection and recovery techniques with respect to various platforms.
- Not only that but also, the paper derived a list of **open research problems** that need to be addressed in future ransomware research and practice.
- Authors believe that this paper will play a crucial role in understanding ransomware research with respect to target platforms and motivating further research.

Thank you!

Any Questions

SEOULTECH
SEOUL NATIONAL UNIVERSITY OF SCIENCE & TECHNOLOGY