Poster: Detecting Ransomware Attacks by Analyzing Replicated Block Snapshots Using Neural Networks

Seok Min Hong
hsmint@hanyang.ac.kr
Hanyang University ERICA
Ansan, Gyeonggi, Republic of Korea

Beom Heyn Kim*
beomheynkim@hanyang.ac.kr
Hanyang University ERICA
Ansan, Gyeonggi, Republic of Korea

Mohammad Mannan
m.mannan@concordia.ca
Concordia University
Montreal, Quebec, Canada

ABSTRACT

Cloud antivirus solutions address limitations of host-based malware detection such as extensive resource consumption. However, they remain vulnerable to sophisticated polymorphic and privileged malware. Also, existing solutions are not suitable to defend against destructive ransomware attacks. We propose an enhancement to existing cloud antivirus solutions that enables deep learning-based block snapshot analysis to detect evasive and privileged ransomware in virtualized environment without requiring any hardware support. Preliminary results validate the proposed approach.

CCS CONCEPTS

• **Security and privacy → Malware and its mitigation**; *Distributed systems security*.

KEYWORDS

Ransomware, Cloud Computing, Antivirus, Distributed Storage Systems, Deep Learning, Virtualization

ACM Reference Format:

Seok Min Hong, Beom Heyn Kim, and Mohammad Mannan. 2024. Poster: Detecting Ransomware Attacks by Analyzing Replicated Block Snapshots Using Neural Networks. In *Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security (CCS '24), October 14–18, 2024, Salt Lake City, UT, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3658644.3691399

1 INTRODUCTION

Cloud antivirus solutions have emerged to address the limitations of host-based malware detection, particularly the significant resource consumption on the host environment [5]. By shifting the resource-intensive task of malware detection to the cloud, these solutions leverage extensive computational power, storage capacity, and real-time threat intelligence, thereby improving device performance and user experience.

Yet, existing cloud antivirus solutions alone may not defend against ransomware, which is gaining popularity among adversaries for financial profits via ransom [10]. Signatures of new threats may not yet be identified in many occasions, or polymorphic malware can successfully evade detection [9]. Additionally, privileged malware can exploit vulnerabilities in operating system kernels, circumventing detection by covertly interfering with the transmission of suspicious files to the cloud [8]. Also, cloud antivirus may not create backups for data files not considered as suspicious. Thus, cloud antivirus may be unable to aid in the restoration of these files following a ransomware attack, rendering user data unavailable.

One notable previous proposal involved replicating block IO at the firmware level of solid-state drives (SSDs), below the operating systems, and sending replicas to remote cloud servers over NVMoF, enabling post-attack data recovery and forensic analysis [7]. However, this proposal does not explore how replicated block IO can be used for virus scanning or the detection of tampered data on the cloud. Additionally, it requires specially designed hardware, which is more expensive to purchase. Furthermore, although Virtual Desktop Infrastructure (VDI) usage is increasing as remote working is becoming the norm, the previous work cannot adapt to a virtual desktop environment where there is no special hardware available for each virtual machine.

Our key observation is that *block-data analysis can always expose sophisticated ransomware attacks.* Because ransomware must tamper with the user data to achieve its goal, all side-effects of its activity can possibly be captured by block snapshots. Thus, our hypothesis is that, by looking at block snapshots changing over time, any sign of attacks mounted by even privileged ransomware can be detected. It is orthogonal to the previous proposals focusing on detecting suspicious I/O behaviours [6], whose limitation is recently shown that ransomware can imitate benign application I/O behaviours and successfully evade state-of-the-art antivirus software [11].

In this work, we propose a novel approach to enhance cloud antivirus solutions to counter evasive and privileged ransomware attacks, without requiring any special hardware. Our proposed solution employs the virtual device layer instead of the physical device layer to cache block snapshots on user devices and periodically replicate these snapshots to cloud servers. Replicated block snapshots then serve as the foundation of a cloud antivirus solution enhanced against ransomware attacks, effectively addressing the limitations previously outlined. We developed a prototype, *RansomSaver* and conducted a preliminary empirical study where we used a deep learning-based classifier to detect any encrypted block snapshot. As a result, we found that it is possible to correctly classify most of block snapshots containing ciphertexts with the high recall score about 95%.

As far as we know, it is the first work analyzing only the block snapshots using deep learning-based classification techniques to detect advanced ransomware attacks. Since *RansomSaver* interposes

*Corresponding Author

CCS '24, October 14–18, 2024, Salt Lake City, UT, USA

ACM ISBN 979-8-4007-0636-3/24/10
https://doi.org/10.1145/3658644.3691399

I/O at the virtual block device layer, no privileged ransomware can hide its activity as long as it encrypts user data. Also, by replicating the block snapshots, our proposal can enable post-attack data recovery and false-positive reduction at the same time.

2 THREAT MODEL

In this work, the target system involves a user device running an operating system environment in a virtual machine (VM). Although it is not restricted, the enterprise environment where VDI is frequently used is the primary target environment. The device can be any computing device capable of running a lightweight virtualization layer and connecting to cloud servers over the network. Examples of such devices include PCs, laptops, tablets, smartphones, and edge servers.

Our solution is not tightly coupled with cloud deployment models. The sole security invariant is to trust the cloud servers. If the cloud servers are protected with trusted computing hardware and users can trust the hardware providers, then a public cloud can be used [3, 4]. Otherwise, the enterprises may operate cloud servers on the on-premise private cloud. Even hybrid cloud can be used by keeping the backup of encrypted block snapshots on the public cloud while analyzing block snapshots on the private cloud. So, we assume adversaries cannot compromise the cloud servers.

On the other hand, adversaries can compromise a user's operating system environment by tricking users into clicking on suspicious links in emails or visiting phishing sites, leading to the installation of powerful malware within the user's VM. However, we assume that adversaries are not able to break out from the VM, because the lightweight virtualization layer has a much smaller trusted computing base (TCB). Note not all user devices can afford hardware-based solutions employing specially designed SSD or trusted computing hardware.

Malware used by adversaries may be polymorphic, with signatures not known in advance. Thus, malware installed on user devices may evade host-based antivirus software. Rootkits are especially dangerous as they can exploit vulnerabilities in the operating system kernels or mount successful privilege escalation attacks. Such privileged malware can defeat host-based antivirus software or agents for cloud antivirus services. Eventually, ransomware may be installed, encrypting the user's data.

3 RANSOMSAVER OVERVIEW

RansomSaver consists of agent and server components that create block snapshots and replicate them from user devices to cloud servers. Figure 1 illustrates the architecture of RansomSaver. In addition to the conventional advantages of cloud antivirus, replicating block snapshots enables several novel techniques that can enhance cloud antivirus.

A RansomSaver agent is built as a block device driver exposed to a VM through the virtualization layer as a passthrough device. The block device driver interposes on any block I/O issued by the VM. For any write to a block, a block snapshot is obtained by updating the previous block content with the interposed write. Additionally, a record of the write operation is appended to a per-VM *write history log*. Subsequently, the block I/O is passed to the underlying physical

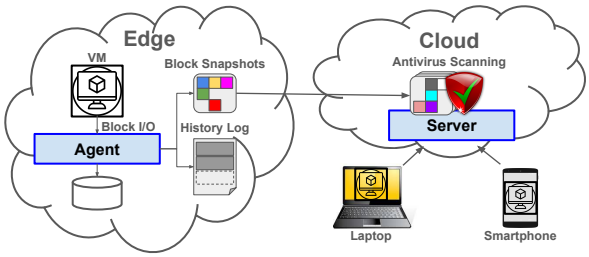

Figure 1: RansomSaver architecture. RansomSaver agent creates block snapshots and periodically replicates to cloud servers where cloud antivirus conducts antivirus scanning. History logs record I/O history leading to the block snapshots for post-attack data recovery. Agents may run on edge server, laptops, smartphones, and so on.

storage device. Block snapshots are periodically transmitted to the cloud servers over the network.

A RansomSaver server runs on a cloud server, accepting incoming block snapshots from an agent on an end device. The RansomSaver server applies block snapshots to block device replicas on cloud servers in a copy-on-write (COW) manner. Every block snapshot of the same epoch from an agent collectively consists a new device snapshot at that epoch.

Multi-Dimensional Scanning. Block snapshots can be used to analyze the state of the block device for signs of ransomware attacks from multiple dimensions. First, block snapshots can be checked for any well-known signatures of malware or ransomware. Second, the entropy of block snapshots can be calculated to measure the probability of encrypted data, aiding in the detection of evasive ransomware attacks based on content. Third, sudden suspicious changes in the state of the block device can be detected by analyzing block snapshots over time. Moreover, this analysis can be performed for each block, for a group of representative blocks, or for the entire block device. Deep learning can be used to identify the surge in the block snapshots from multiple dimensions.

False-Positive Reduction. Suspicious block snapshots detected using machine learning techniques may often result in false positives [1]. Maintaining block device replicas by applying block snapshots to mirror those on end devices allows us to reduce false positives through additional analysis on cloud servers. More specifically, RansomSaver can mount a replicated block device on a cloud server's file system and conduct further analysis of each suspicious file by leveraging high-level information. For instance, if a file name ends with .pdf or .docx file extension, it should not contain any encrypted text when rendered by a PDF viewer or a word processor.

Distributed Scanning and Analysis. Speeding up attack detection is desirable, but scanning a large amount of block data can take a long time. The advantage of cloud antivirus is the ability to flexibly expand resource utilization over a large scale of underutilized cloud servers on demand. Block snapshots can be partitioned and distributed to multiple cloud servers via distributed storage systems such as DynamoDB, Cassandra, MongoDB, or Redis. Then, analysis can be performed in parallel on distributed cloud servers.

Table 1: Performance of RansomSaver For Classifying Ciphertexts and Plaintexts Based on Block Snapshot Analysis.

Ransomware Sample	Accuracy (%)	Precision Score (%)	Recall Score (%)	F1 Score (%)
WannaCry	77.92	64.23	95.20	76.71
Animagus	72.56	40.80	94.87	57.06

Similarly, false-positive reduction can be conducted on many distributed cloud servers in parallel as well.

4 EMPIRICAL STUDIES

We performed preliminary empirical exploration on RansomSaver, and found that analyzing block snapshots using deep learning-based classification techniques can differentiate ciphertexts from plaintexts with the low false negative rate—i.e., failing to detect ciphertexts.

Evaluation Setup. RansomSaver is implemented on top of Rocky [2]. A RansomSaver agent generates block snapshots and transmits those to a RansomSaver server, which stores block snapshots in a DynamoDB instance. Block snapshots stored in the DynamoDB instance are used to train a neural network model.

Our experimentation was conducted on a desktop PC in our lab equipped with an Intel i5 processor, 16 GB RAM, and 512 GB NVMe SSD. The host OS was Ubuntu 22.04, and we used VirtualBox to run a VM installed with a guest OS, Windows 10 Home Edition. The VM serves a contained environment to run ransomware samples.

We collected the initial dataset by downloading text data from the internet, which generates non-encrypted block snapshots. Additionally, we compressed the text data to create additional non-encrypted block snapshots containing compressed data. Furthermore, we obtained encrypted block snapshots by running Animagus or WannaCry in the VM. As a result, we obtained 5316 plaintext blocks, 3280 compressed blocks, 2043 ciphertext blocks from Animagus, and 5311 ciphertext blocks from WannaCry. Among these, we used 80% for training and 20% for inference.

For the training datasets, non-encrypted and encrypted block snapshots were labeled. Then, test datasets were used to evaluate the classification capability of the trained model. The trained neural network had 1,227,777 parameters, trained with a batch size of 64 for 70 steps taking one minute and thirty seconds. We used ResNet as a base model with a sigmoid activation function for the final layer and rectified linear activation for other layers. We used Binary Cross Entropy (BCELoss) to calculate the loss and optimized using Adam with a learning rate of 0.001.

Evaluation Results. Evaluation results are summarized in Table 1. The recall score is about 95% for both Animagus and WannaCry, which implies that our model rarely misses possible ransomware attacks—i.e., it rarely misclassifies ciphertexts as plaintexts. Thus, we claim that applying deep learning-based classification techniques to detect Ransomware attacks can be effective.

We observed notable differences in precision and F1-score. The reason is the imbalanced size of test datasets between WannaCry and Animagus. This results in 1011 and 388 true positives for WannaCry and Animagus, respectively, while 563 false positives are observed for both. We envision the sensitivity of the precision score will be masked out, as we collect more datasets in the future.

Moreover, accuracy, precision, and F1-score will be further improved by reducing false positives and false negatives by using much larger datasets along with validation sets and by increasing the number of parameters. Additionally, we will explore various ways of analyzing block snapshots over multiple epochs.

5 CONCLUSION

In this study, we introduced RansomSaver, a novel approach to enhancing cloud antivirus solutions to effectively counter evasive and privileged ransomware attacks without requiring specialized hardware. Unlike previous proposals relying on I/O patterns, our solution enables the analysis of block snapshots using neural networks. Based on the preliminary results, we envision that using neural networks can be effectively applied to classify block snapshots, offering a powerful tool in the fight against imitation-based and privileged ransomware attacks.

6 ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (No. RS-2023-00244368).

REFERENCES

[1] Robert Bold, Haider Al-Khateeb, and Nikolaos Ersotelos. 2022. Reducing False Negatives in Ransomware Detection: A Critical Evaluation of Machine Learning Algorithms. *Applied Sciences* 12 (12 2022), 12941. https://doi.org/10.3390/app122412941

[2] Beom Heyn Kim and Hyoungshick Kim. 2021. Rocky: Replicating Block Devices for Tamper and Failure Resistant Edge-based Virtualized Desktop Infrastructure. In *Proceedings of the 37th Annual Computer Security Applications Conference* (Virtual Event, USA) *(ACSAC '21)*. Association for Computing Machinery, New York, NY, USA, 285–296. https://doi.org/10.1145/3485832.3485886

[3] Klaudia Krawiecka, Arseny Kurnikov, Andrew Paverd, Mohammad Mannan, and N. Asokan. 2018. SafeKeeper: Protecting Web Passwords using Trusted Execution Environments. *WWW '18: Proceedings of the 2018 World Wide Web Conference*, 349–358. https://doi.org/10.1145/3178876.3186101

[4] Arseny Kurnikov, Andrew Paverd, Mohammad Mannan, and N. Asokan. 2018. Keys in the Clouds: Auditable Multi-device Access to Cryptographic Credentials. In *Proceedings of the 13th International Conference on Availability, Reliability and Security* (Hamburg, Germany) *(ARES '18)*. Association for Computing Machinery, New York, NY, USA, Article 40, 10 pages. https://doi.org/10.1145/3230833.3234518

[5] Jon Oberheide, Evan Cooke, and Farnam Jahanian. 2008. CloudAV: N-Version Antivirus in the Network Cloud. In *USENIX Security Symposium*. 91–106.

[6] Harun Oz, Ahmet Aris, Albert Levi, and A. Selcuk Uluagac. 2022. A Survey on Ransomware: Evolution, Taxonomy, and Defense Solutions. *ACM Comput. Surv.* 54, 11s, Article 238 (sep 2022), 37 pages. https://doi.org/10.1145/3514229

[7] Benjamin Reidys, Peng Liu, and Jian Huang. 2022. RSSD: defend against ransomware with hardware-isolated network-storage codesign and post-attack analysis. In *Proceedings of the 27th ACM International Conference on Architectural Support for Programming Languages and Operating Systems* (Lausanne, Switzerland) *(ASPLOS '22)*. Association for Computing Machinery, New York, NY, USA, 726–739. https://doi.org/10.1145/3503222.3507773

[8] Symantec Threat Hunter Team. 2024. Ransomware Attackers May Have Used Privilege Escalation Vulnerability as Zero-day. https://symantec-enterprise-blogs.security.com/threat-intelligence/black-basta-ransomware-zero-day (last accessed: 30/06/2024).

[9] The BlackBerry Cylance Threat Research Team. 2019. Threat Spotlight: Virlock Polymorphic Ransomware. https://blogs.blackberry.com/en/2019/07/threat-spotlight-virlock-polymorphic-ransomware (last accessed: 30/06/2024).

[10] Kevin Townsend. 2024. The Ransomware Threat in 2024 is Growing: Report. https://www.securityweek.com/the-ransomware-threat-in-2024-is-growing-report/ (last accessed: 30/06/2024).

[11] Chijin Zhou, Lihua Guo, Yiwei Hou, Zhenya Ma, Quan Zhang, Mingzhe Wang, Zhe Liu, and Yu Jiang. 2023. Limits of I/O Based Ransomware Detection: An Imitation Based Attack. In *2023 IEEE Symposium on Security and Privacy (SP)*. 2584–2601. https://doi.org/10.1109/SP46215.2023.10179372