Amoeba: An Autonomous Backup and Recovery SSD for Ransomware Attack Defense

Donghyun Min\textsuperscript{1}, Donggyu Park\textsuperscript{1}, Jinwoo Ahn\textsuperscript{1}, Ryan Walker\textsuperscript{2}, Junghee Lee\textsuperscript{2}, Sungyong Park\textsuperscript{1}, Youngjae Kim\textsuperscript{1}

Presenter: Donghyun Min
Feb 19, 2019 @ HPCA’19
Ransomware

User -> Encryption

Encryption

Ransomware
Ransomware

Huge Financial Loss

Encryption

Ransom fee
Damage of Ransomware Attack

- Many areas are suffering from damage
  - Public institutions
  - Government, industry

Consider the following example:
Buffalo, NY, last July estimated

Ransomware-related damage cost will reach $20 billion by 2021!
How to Defend against Ransomware Attack

- Backup method

Original copy

Backup copy
How to Defend against Ransomware Attack

- Approach 1: Host-level backup
  - Backup on Local File system
  - Backup on Remote machine

- Approach 2: Device-level backup
  - FlashGuard [CCS’17]
  - SSD-Insider [ICDCS’18]
  - Amoeba [CAL’18]
Approach 1: Host-level Backup

- Backup inside File system
  - Extra storage space is required.
  - Ransomware with kernel privilege can disable backup process.

- Backup on Remote machine

1. Extra storage space is required.
2. Ransomware with kernel privilege can disable backup process.
Approach 2: Device-level Backup

- FlashGuard [CCS’17]
- SSD-Insider [ICDCS’18]
Opportunities: Out-of-Place Update in an SSD

Flash Translation Layer (FTL)

Address Translation (LPN, PPN)

... (10, 2) (20, 3) ...

SSD

NAND Flash memory

VALID

VALID

VALID

VALID

VALID

VALID

VALID

VALID

VALID

Physical Block

Physical Page
Opportunities: Out-of-Place Update in an SSD

Encrypt File(A) by Ransomware

Flash Translation Layer (FTL)

SSD

NAND Flash memory

 VALID
 VALID
 VALID
 VALID
 VALID
 VALID
 VALID
 VALID
 VALID

Address Translation (LPN, PPN)

LPN 10

(10, 2)
(20, 3)

In-place Update

Encrypt File(A) by Ransomware

Opportunities: Out-of-Place Update in an SSD

Encrypt File(A) by Ransomware

Flash Translation Layer (FTL)

SSD

NAND Flash memory

 VALID
 VALID
 VALID
 VALID
 VALID
 VALID
 VALID
 VALID
 VALID

Address Translation (LPN, PPN)

LPN 10

(10, 2)
(20, 3)

In-place Update

Encrypt File(A) by Ransomware
Opportunities: Out-of-Place Update in an SSD

Encrypt File(A) by Ransomware

SSD

Address Translation (LPN, PPN)

Flash Translation Layer (FTL)

NAND Flash memory

Out-of-place Update

LPN 10

(10, 4)

(20, 3)

VALID

INVALID

VALID

VALID

VALID

VALID

VALID

VALID
Opportunities: Out-of-Place Update in an SSD

Encrypt File(A) by Ransomware

Invalid page is actually an original page for recovery.

1. We can save storage space for backup because **additional backup space is not required**.
2. Device-level backup can become more secure because **backup copy cannot be seen from ransomware application**.
Challenges

Encrypt File(A) by Ransomware

Overwrites on File(B) by Normal User

SSD

Address Translation (LPN, PPN)

VALID
INVALID
VALID
VALID
VALID
VALID
VALID
VALID
VALID
VALID

Out-of-place Update

Flash Translation Layer (FTL)

NAND Flash memory

Backup
Challenges

Encrypt File(A) by Ransomware

Overwrites on File(B) by Normal User

- Flash Translation Layer (FTL)
- NAND Flash Memory
- SSD

Address Translation (LPN, PPN)

- (10, 4)
- (20, 3)

Out-of-place Update

LPN 20
SSD should keep invalid pages as backup only for updates by ransomware.
Summary: Limitations of Previous Works [CCS’17, ICDCS’18]

1. Lack of accurate ransomware detection algorithms
   - Detection solely relies on I/O access pattern (e.g., Write Intensity)
     - **False Positive (FP)**
     - **GC overhead**
     - **False Negative (FN)**
     - **Recovery failure**
Summary: Limitations of Previous Works [CCS’17, ICDCS’18]

1. Lack of accurate ransomware detection algorithms
   - Detection solely relies on I/O access pattern (e.g., Write Intensity)
     ➔ *False Positive (FP)* ➔ GC overhead
     ➔ *False Negative (FN)* ➔ Recovery failure

2. High unnecessary space overhead due to lack of intelligent backup mechanisms
   - Unnecessary backup pages increase GC overhead.
Our Approach [Amoeba, CAL’18]

1. We use a **content-based detection** technique for high ransomware detection rate.

2. We implement an **intelligent backup management mechanism** to minimize space overhead for backup pages.
Challenge 1: How to Apply Content-based Detection at High Speed

- Content-based detection offers high ransomware detection rate, but, it is highly time-consuming because it requires huge computation power for old and new comparison for similarity and entropy computation.
Challenge 2: How Accurately Detect Ransomware Attack

- Ransomware detection algorithm needs to be developed by considering three indicators all together should be required for high detection rate.

![Diagram showing I/O access pattern, Write Intensity, Similarity (A and A'), and Entropy.](image)
Challenge 2: How Accurately Detect Ransomware Attack

- If only **Write Intensity** is used, it often misjudge normal requests and ransomware attacks.

- If only **Similarity** and **Entropy** are used, it cannot distinguish legitimate encryption applications using compression and PGP cryptographic library from ransomware attacks.
Challenge 3: How to Minimize Backup Space Overhead

We should be able to identify only necessary backup pages for recovery among backup pages.
Amoeba:
An Autonomous Backup and Recovery SSD for Ransomware Attack Defense
Amoeba System Architecture
Amoeba System Architecture

- Amoeba DMA

Host machine

SSD

DRAM Buffer

SSD Controller

Flash Translation Layer (FTL)

DRAM Controller

Amoeba DMA

Flash Controller

NAND Flash

15
Amoeba System Architecture

- Ransomware Attack Risk Indicator (RARI)
Amoeba System Architecture

- Intelligent Backup Mechanism

![Diagram of Amoeba System Architecture](image-url)
1. Amoeba DMA Engine

- Amoeba DMA engine for computing similarity, entropy
1. Amoeba DMA Engine

- Amoeba DMA engine for computing similarity, entropy
  - Basic DMA (Existing DMA)
1. Amoeba DMA Engine

- Amoeba DMA engine for computing similarity, entropy
  - Basic DMA (Existing DMA)
  - Amoeba DMA
1. Amoeba DMA Engine

- Amoeba DMA engine for computing similarity, entropy
  - Basic DMA (Existing DMA)
  - Amoeba DMA

Calculation delay can be hidden.
2. Ransomware Attack Risk Indicator (RARI)

- We establish a model that combines three indicators (write intensity, similarity, and entropy) to form a RARI value.
2. Ransomware Attack Risk Indicator (RARI)

- We establish a model that combines three indicators (write intensity, similarity, and entropy) to form a RARI value.
2. Ransomware Attack Risk Indicator (RARI)

- We establish a model that combines three indicators (write intensity, similarity, and entropy) to form a RARI value.

\[ \alpha \times SIM + \beta \times ENT + \gamma \times INT + \delta \]

Threshold (fixed) = 0.5f

Probability
3. Intelligent Backup Control Mechanism

- We can accurately detect backup pages using RARI values. Thus, we can only maintain a backup page per logical page.

We can completely go away unnecessary backup pages in an SSD.
3. Intelligent Backup Control Mechanism

- Recovery Procedure

Recovery request

![Diagram showing SSD NAND Block recovery process]
Evaluation Methodology

- MSR Disksim SSD Simulator
- Workload
  • Linux Erebus ransomware
  • User’s normal I/O
- Simulation setup
  • SSD Occupancy: 20%, 40%, 80%
  • Page Size: 8 KB, # of page per block: 128
- Comparison
  • Baseline: SSD without backup mechanism
  • FlashGuard
  • SSD-Insider
  • Amoeba
Result 1: Average Response Time

- Baseline
- FlashGuard[1]
- FlashGuard[2]
- FlashGuard[4]
- FlashGuard[8]
- SSD-Insider
- Amoeba

Normalized Avg. Response Time (ms)

- SSD page occupancy 20%
- SSD page occupancy 40%
- SSD page occupancy 80%
Result 1: Average Response Time
Result 1: Average Response Time

In worst case, response time of Amoeba only increased by 8% compared to baseline.
## Result 2: Detection Accuracy

<table>
<thead>
<tr>
<th>Detection Method</th>
<th>Number of Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>FlashGuard</td>
<td>70000</td>
</tr>
<tr>
<td>SSD-Insider</td>
<td>11.11%</td>
</tr>
<tr>
<td>Amoeba</td>
<td>2.79%</td>
</tr>
</tbody>
</table>

Amoeba has only less than 1% false detection.
Conclusion

• We presented **Amoeba: An Autonomous Backup and Recovery SSD for Ransomware Attack Defense.**
  • Implemented **Amoeba DMA Hardware engine** to compute content-based detection algorithm.
  • Proposed a **Ransomware Attack Risk Indicator (RARI) metric.**
  • Provided **Intelligent Backup and Recovery mechanism.**
Thank you

Q & A

Donghyun Min
mdh38112@sogang.ac.kr
Sogang University, South Korea
Backup slides 1: GC Calls

Number of GC Calls

- SSD page occupancy 20%
- SSD page occupancy 40%
- SSD page occupancy 80%

- Baseline
- FlashGuard[1]
- FlashGuard[2]
- FlashGuard[4]
- FlashGuard[8]
- SSD-Insider
- Amoeba