APRON: Authenticated and Progressive System Image Renovation

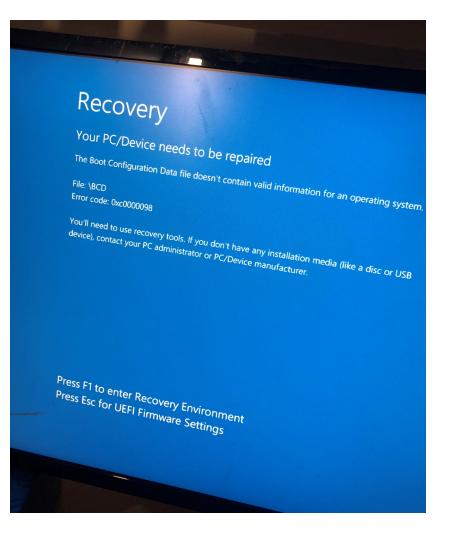
Sangho Lee

2023 USENIX Annual Technical Conference July 10, 2023



Devices or systems might be corrupt





* https://www.wirelesshack.org/fixes-for-your-device-is-corrupted-and-cannot-be-trusted.html

* https://answers.microsoft.com/en-us/windows/forum/all/recovery-your-pcdevice-needs-to-be-repaired/6d8a33b0-ccdb-43f7-b50c-1937aec79033

Why and how are systems corrupt?

6	Wana Decrypt0r 2.0		
	Ooops, your files have been encrypted! English What Happened to My Computer? Your important files are encrypted. Many of your documents, photos, videos, databases and other files are no longer accessible because they have been encrypted. Maybe you are busy looking for a way to	<pre>////////////////////////////////////</pre>	
Payment will be raised on 5/16/2017 00:47:55 Time Left	recover your files, but do not waste your time. Nobody can recover your files without our decryption service. Can I Recover My Files? Sure. We guarantee that you can recover all your files safely and easily. But you have not so enough time. You can decrypt some of your files for free. Try now by clicking <decrypt>.</decrypt>	134.9180682] Buffer I/0 error on dev sdal, logical block 0, async page read fsck.ext4[134.918208] sd 0:0:0:0: [sda] tag#0 FAILED Result: hostbyte=DID_BAU_ TARGET driverbyte=DRIVER_OK : Attempt to read block from filesystem resulted in short read while trying to r e-open /dev/sdal[134.918263] sd 0:0:0:0: [sda] tag#0 CDB: Synchronize Cache(10) 35 00 00 00 00 00 00 00 00 00 [134.918266] blk_update_request: I/0 error, dev sda, sector 0	Windows Update You're up to date Last checked: Today, 6:50 AM Check for updates
82:23:57:37	But if you want to decrypt all your files, you need to pay. You only have 3 days to submit the payment. After that the price will be doubled. Also, if you don't pay in 7 days, you won't be able to recover your files forever. We will have free events for users who are so poor that they couldn't pay in 6 months. How Do I Pay? Payment is accepted in Bitcoin only. For more information, click <about bitcoin="">.</about>	/dev/sda1: ********** WARNING: Filesystem still has errors ********* fsck exited with status code 12 done.	More options Get the latest updates as soon as they're available Pause for
Time Left 06:23:57:37	Please check the current price of Bitcoin and buy some bitcoins. For more information, click «How to buy bitcoins». And send the correct amount to the address specified in this window. After your payment, click «Check Payment». Best time to check: 9:00am - 11:00am CMT Bear Mardense Pedera Send \$300 worth of bitcoin to this address:	Failure: File system check of the root filesystem failed The root filesystem on /dev/sda1 requires a manual fsck BusyBox v1.22.1 (Ubuntu 1:1.22.0-19ubuntu2) built-in shell (ash) Enter 'help' for a list of built-in commands.	Inis setting isn't available due to your organization's policy Learn more 100 years Image: Description of the setting isn't available due to your organization's policy Pause for 1 week
<u>About biccoin</u> <u>How to buy bilcoins?</u> <u>Contact Us</u>	Ditcoin ACCEPTED HERE 12t9YDPgwueZ9NyMgw519p7AA8isjr6SMw Copy Check Payment Decrypt	(initramfs)	

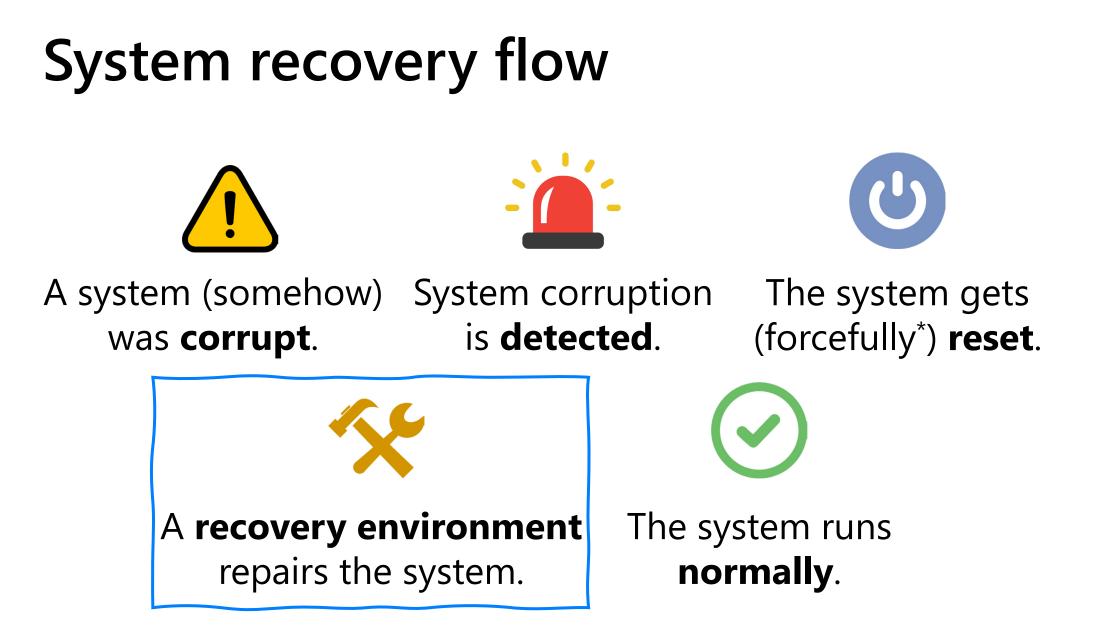
Attacks/Malware

Software/hardware errors

Postponed updates

* https://securelist.com/wannacry-ransomware-used-in-widespread-attacks-all-over-the-world/78351/

* https://askubuntu.com/questions/972978/fsck-reports-that-filesystem-still-has-errors/



* M. Xu et al., "Dominance as a New Trusted Computing Primitive for the Internet of Things," IEEE S&P 2019.

What does a recovery environment do?

- Prepare a separate environment with recovery tools
- Completely recover corrupt system storage using a reference image
 - Download one from a remote location
 - Use one stored in a safe local location
- Verify the recovered system and restart it

Observation: Recovery time \approx system downtime

- Recovery environment is designed to be minimal.
 - Contain recovery-related tools only
 - Lack everyday programs, libraries, ...
- System is effectively **down** during recovery.
- Delaying recovery lengthens system downtime.
 - Download a large reference image file (and decompress it)
 - Reimage the entire storage and verify it

Questionable ways to speed up recovery

- Use a stored system image (e.g., A/B partition)
 Stored image can also be corrupted or outdated.
- Selectively fix corrupted files/blocks (delta recovery)
 - Difference calculation (e.g., rdiff) and scattered disk updates are slow.

Key idea: Progressive recovery

Defer the recovery of data blocks until they are needed

- A complete system image is not needed to start it.
 - Boot: A small part of it (i.e., kernel, initramfs) is required.
 - Execution: Some parts related to active tasks are required.

Selectively/partially recover the system and make it progressively recover the remainder on demand



Key idea: Authenticated recovery

Verify every storage access to detect/recover invalid blocks

- Apps must be able to get valid data when they access system storage even if we defer its recovery.
 - Accessing invalid data blocks is prohibited.
- >Interpose storage access to always return valid data
 - Verify a requested block with authenticated metadata
 - Serve the block after recovery if it was corrupt or outdated

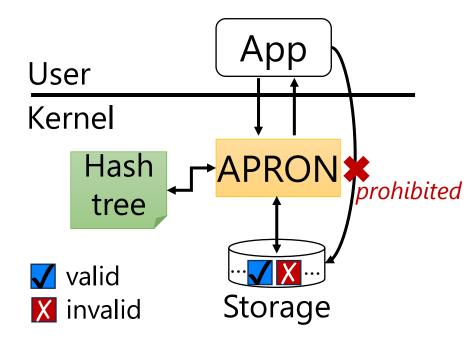
Image-based (immutable) operating system

- Management server builds an up-to-date system image and provisions it to managed devices.
- Operating system kernel enforces the read-only property of the system image.
- Store and manage mutable configuration and data in a separate place

Preparation and device initialization

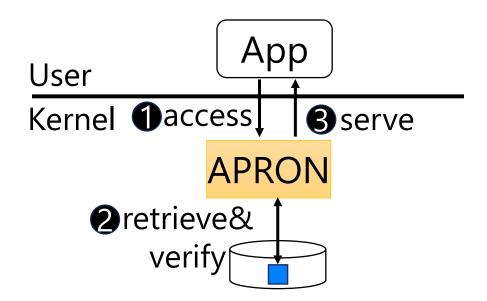
- Server
 - Calculate a hash tree over the latest system image
 - Sign its root hash concatenated with a version number
 - Serve the system image and signed metadata
- Device (initramfs)
 - Download and verify the latest metadata
 - Create a storage layer over system storage with the metadata

APRON storage layer

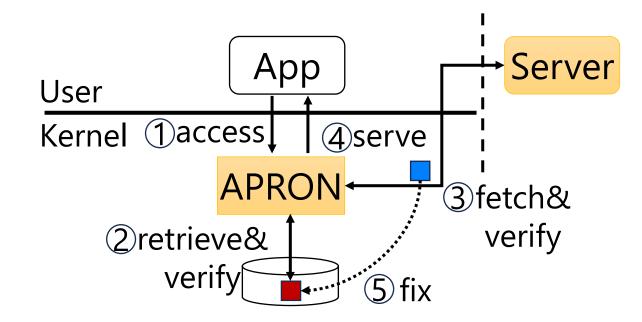


- Intervene with every access to system storage
- Verify each requested block using the hash tree
- Repair requested blocks if they are invalid (hash mismatch)

On-demand renovation (valid block)

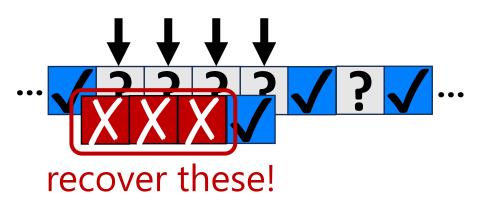


On-demand renovation (invalid block)



Background prefetcher

- Renovate non-yet requested blocks in background
 - Eventually recover the entire system storage
 - Wake up if there is no in-flight storage access
- Find and batch repair consecutive invalid blocks
 - Inspect unidentified blocks until it encounters a valid block
 - Recover the found invalid blocks together (for throughput)

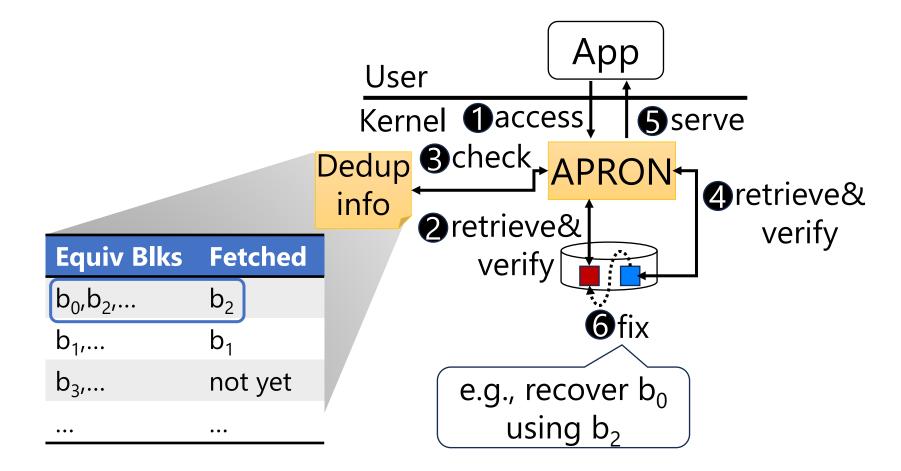


unidentified
valid
invalid

In-place renovation with deduplication

- Fix requested blocks with equivalent blocks in storage
- Rely on static and dynamic deduplication information
 - Server pre-computes sets of equivalent blocks.
 - Device tracks whether it has fetched any block of each set.
- Fetch a remote block for recovery only if
 - It is unique; or
 - None of its equivalent block has been fetched.

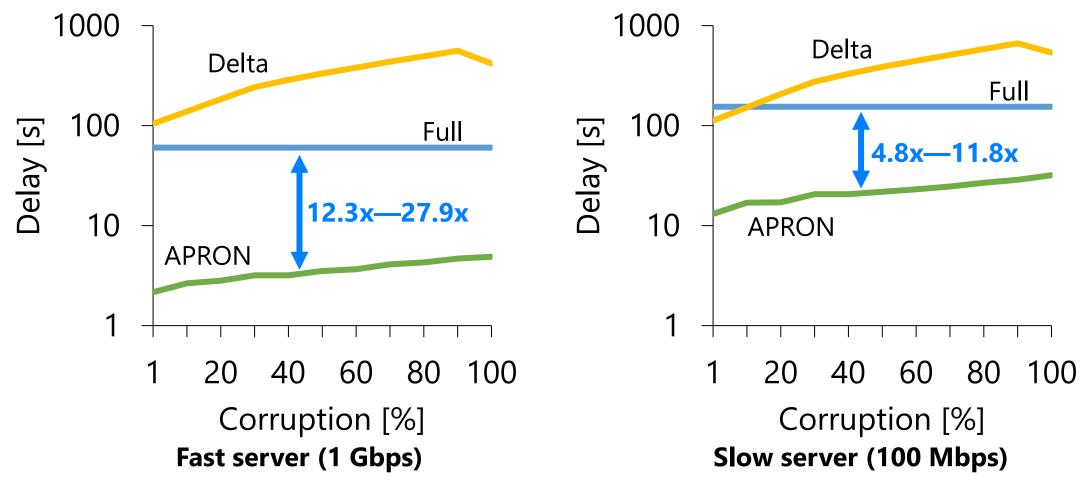
In-place renovation (invalid block)



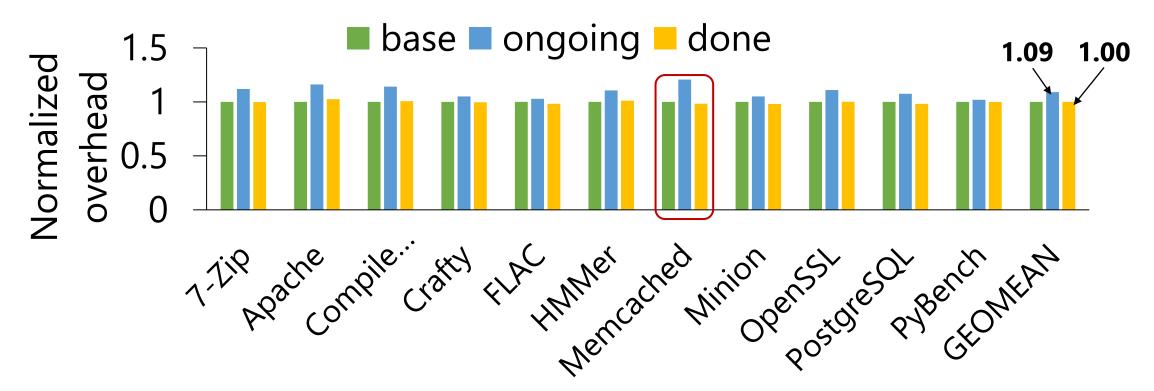
Evaluation setup

- Machineries
 - Client: Desktop CPU (six cores) & PCIe NVMe SSD
 - Fast-network server (1 Gbps): Desktop CPU (four cores) & SATA SSD
 - Slow-network server (100 Mbps): Server VCPU (two cores) & virtual SSD
- System image
 - 10 GiB of Ubuntu Server 20.04 installation
 - Randomly corrupt 1%—100% of it

System downtime

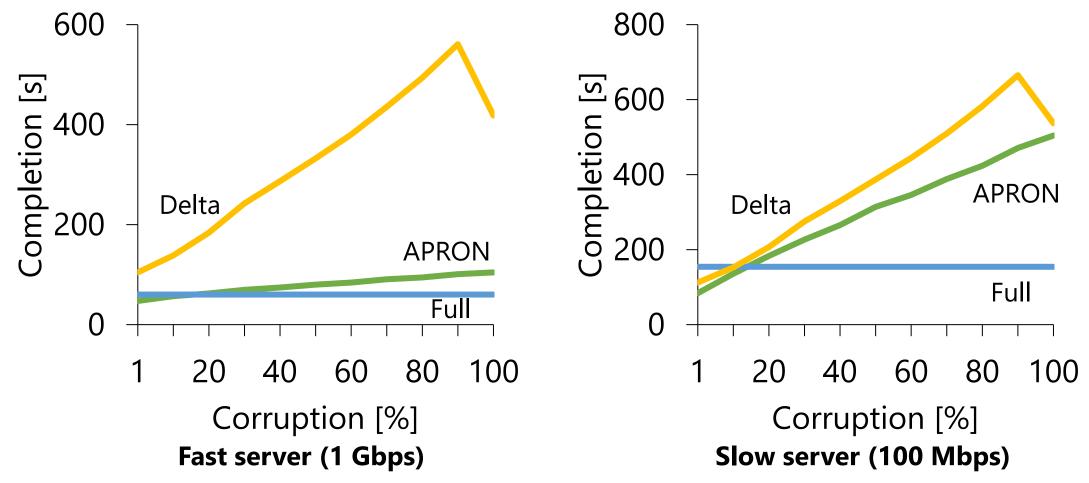


Runtime overhead: Phoronix test suite



- Fully corrupted system storage and slow-network server
- Affect I/O-intensive workloads (e.g., Memcached: 21%)

Complete renovation time



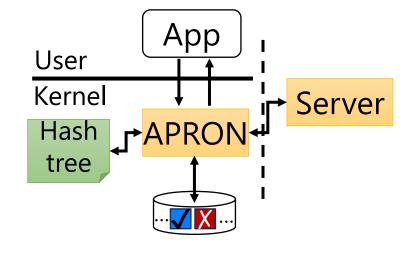
Conclusion

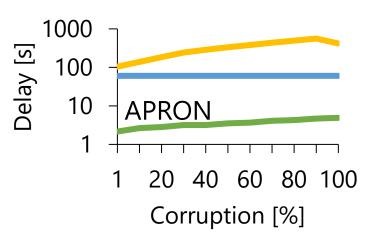
Slow recovery makes the system unavailable.

Securely defer recovery for instant system availability

Ensure short downtime and low runtime overhead







Sangho.Lee@microsoft.com <u>https://github.com/microsoft/APRON</u> (soon)