Zero-Click, Old Tricks: Anatomy of the 2025 WhatsApp–ImageIO Exploit Chain

WAGNER Ştefan-Daniel
OLTEAN Dan-Gabriel MATVEEV Victor-Nicolae
Facultatea de Ştiinţe Aplicate
Teoria Codării şi Stocării Informaţiei
Universitatea Naţională de Ştiinţă şi Tehnologie POLITEHNICA Bucureşti (UNSTPB)
Coordonator: Emil Simion
Semestrul II, 2025–2026

***Abstract*—Zero-click exploits (attacks that compromise a device without any user interaction) represent the most asymmetric threat in mobile security. This paper analyzes the 2025 WhatsApp–ImageIO zero-click exploit chain, which combined a WhatsApp linked-device authorization bypass (CVE-2025-55177, CWE-863) with a heap out-of-bounds write in Apple's ImageIO framework (CVE-2025-43300, CWE-787) to achieve remote code execution on iOS devices, targeting approximately 200 journalists and civil-society figures over 90 days. We reconstruct the chain from public primary sources (vendor advisories, patch diffs, and independent reverse engineering) and use the Exploit Education Phoenix `heap-two` exercise as a pedagogical bridge to demonstrate that the core exploitation primitive (heap overflow into an adjacent function pointer) is identical at the textbook and production levels. Comparing the 2025 chain with the structurally identical FORCEDENTRY chain (2021, CVE-2021-30860), we argue that the "delivery primitive plus image parser memory corruption" pattern is inherent to the current architecture of mobile media processing and that the deployed mitigation stack (ASLR, sandboxing, pointer authentication) raises the cost of exploitation without eliminating the root cause. We recommend incremental rewriting of high-exposure C/C++ parsers in memory-safe languages and deployment of hardware memory tagging as structural countermeasures.**

I. Introduction

In August 2025, Apple and WhatsApp released emergency advisories for CVE-2025-43300 and CVE-2025-55177, a pair of vulnerabilities that, chained together, had already been used in the wild to compromise the mobile devices of fewer than 200 users over the preceding 90 days [1], [2]. Independent reporting identified many of the targets as journalists, human-rights defenders, and civil-society figures. The chain required no clicks, no downloads, and no error on the victim's part: receiving a WhatsApp message was sufficient. Both CVEs were added within weeks to the U.S. Cybersecurity and Infrastructure Security Agency (CISA) Known Exploited Vulnerabilities (KEV) catalog [3], mandating remediation across federal networks.

Zero-click chains of this shape have become the defining offensive capability of commercial spyware. Unlike phishing, which relies on social engineering, and unlike drive-by compromise, which relies on the victim visiting an attacker-controlled URL, a zero-click chain weaponizes the infrastructure that modern mobile operating systems use to render inbound media. Knowledge of the victim's phone number is sufficient; no further action on the victim's part is required. The asymmetry of this attack model (bounded defender effort against unbounded attacker opportunity) makes zero-click delivery the most consequential primitive in targeted surveillance today [4].

What makes the 2025 chain technically striking is its genealogy. Both vulnerabilities fire in code written in C, a language first standardized in 1989 whose memory model (Section II) provides no enforcement of bounds, type safety, or lifetimes. The exploitation stage is a heap out-of-bounds write in an image parser, a bug class first formalized by Aleph One in 1996 [5] and still ranked first in the MITRE CWE Top 25 list three decades later [6]. Despite repeated calls from CISA and the U.S. Office of the National Cyber Director for migration to memory-safe languages [7], the media-parsing frameworks at the core of every major mobile platform remain overwhelmingly C and C++. The result is a structural mismatch: a half-century-old memory model is entrusted with the first parsing step of every image delivered to every networked device.

This is not the first time the pattern has produced high-profile civilian casualties. In 2021, Citizen Lab and Google Project Zero reverse-engineered FORCEDENTRY (CVE-2021-30860), an NSO Group Pegasus zero-click chain with nearly identical architecture: iMessage delivery, CoreGraphics image parsing, heap corruption [8], [9]. Four years later, the 2025 WhatsApp–ImageIO chain reproduces the same three-step pattern in a different ecosystem, a different messenger, a different image format, and a different parser, with the same outcome. This recurrence across independent platforms and independent vendor teams is the central empirical observation motivating this

study.

Research question. This paper asks: *how do decades-old memory-corruption patterns in C enable modern zero-click exploit chains, and what does the 2025 WhatsApp–ImageIO attack reveal about the persistence of these vulnerabilities in industry-critical software?*

Contributions. This paper makes three contributions:

1) A reconstruction of the 2025 zero-click chain from public primary sources (vendor advisories, patch diffs, and independent reverse engineering [10]), covering the delivery primitive (CVE-2025-55177, CWE-863) and the exploitation primitive (CVE-2025-43300, CWE-787), and showing how the two stages compose into a complete, interaction-free compromise.
2) A pedagogical bridge from the Exploit Education Phoenix `heap-two` exercise to CVE-2025-43300, demonstrating that the core mechanic (a heap buffer overflowed into adjacent function-pointer-bearing data) is the same primitive at two drastically different complexity levels. The bridge is intended for readers whose prior exposure to memory corruption stops at textbook stack smashing.
3) A structural argument that the FORCEDENTRY → WhatsApp–ImageIO lineage, spanning four years and two independent ecosystems, is evidence that the current mitigation stack (ASLR, sandboxing, pointer authentication) has not closed the fundamental attack class; it only raised its price.

Scope and non-goals. This paper does not present a new exploit, a new vulnerability, or proprietary reverse engineering. All technical material is drawn from publicly disclosed sources. The Phoenix `heap-two` exercise is treated as a controlled pedagogical artifact, not a substitute for the real chain.

Roadmap. Section II introduces the C memory model, the two CWE classes at play, the iOS media pipeline, and the FORCEDENTRY precedent. Section III describes our CVE selection criteria, analysis framework, and Phoenix setup. Section IV is the core technical analysis of the 2025 chain. Section V reflects on why the current mitigation stack fails to close this attack class. Section VI concludes.

II. Background

This section establishes the minimum technical vocabulary needed to follow the two-vulnerability chain dissected in the case study. Readers are assumed to be computer-science graduates without a specialization in systems security; every term is introduced the first time it is used. We cover, in order, the C memory model (II-A), the two weakness classes that appear in the chain (II-B), the iOS media-processing pipeline in which both vulnerabilities fire (II-C), the mitigations that are supposed to prevent exploitation (II-D), and the 2021 FORCEDENTRY incident which is the historical precedent for the 2025 attack (II-E).

A. The C Memory Model: A Brief Tour

C is a language that trades safety for control. Unlike managed languages such as Java, Python, or Swift, C gives the programmer direct access to raw memory through pointers, with no runtime enforcement of object bounds, type safety, or lifetimes. A running C program organizes its memory into several regions; two are central to exploitation. The *stack* is a last-in-first-out region where local variables and function call frames live. Every function call pushes a new frame containing parameters, locals, and the saved return address: the instruction to jump to after the callee returns. The *heap* is a region for dynamically allocated memory (via `malloc`, `calloc`, `new`). Blocks are allocated and freed explicitly by the programmer; the heap allocator maintains bookkeeping metadata (block size, free/used flags, chunk pointers) interleaved with or adjacent to user data.

A pointer in C is simply an integer that names a memory address. Dereferencing a pointer (`*p` for read, `*p = v` for write) performs raw memory access with no check that `p` points to a valid, in-bounds location of the expected type. This is the root of the entire memory-corruption family: if an attacker can drive a write slightly past where it was supposed to land, they can alter state (other variables, function pointers, return addresses, allocator metadata) that the program will later trust and dereference. Aleph One's 1996 essay "Smashing the Stack for Fun and Profit" [5] first formalized the stack-based version of this idea; three decades later, the same pattern continues to produce exploitable bugs at the top of the MITRE CWE Top 25 list [6]. The heap-specific variant of this idea was formalised in 2001 by the Phrack article "Once upon a free()" [11], which showed how `dlmalloc`'s bookkeeping metadata (stored adjacent to user data in the same heap region) could be weaponised into a write-anywhere primitive via the allocator's own unlink routine; three decades later, the structural insight remains valid. Listing 1 shows the minimal C pattern that produces this class of bug.

Listing 1. Minimal heap out-of-bounds write. If `len > 64`, bytes spill from `buf` into adjacent heap memory, corrupting allocator metadata or attacker-reachable data (CWE-787).

```
char *buf = malloc(64);  /* allocate 64-byte
    heap chunk */

/* No bounds check: if len > 64, bytes at buf
    [64..len-1]
   are written into the *next* heap chunk.
                    */
memcpy(buf, attacker_input, len);

/* The corruption is invisible here -- it
    surfaces later
   when the program dereferences overwritten
       state.     */
```

B. Vulnerability Taxonomy

Two CWE classes drive the attack chain studied in this work.

CWE-787: Out-of-bounds write [6] is the class in which the program writes past the intended end of a buffer. On the heap, an out-of-bounds write may corrupt adjacent allocations or allocator bookkeeping; on the stack, it may corrupt saved registers or the return address. CWE-787 has held the #1 position in the MITRE Top 25 Most Dangerous Software Weaknesses for multiple consecutive years and accounts for the majority of critical remote-code-execution bugs patched in commodity operating systems [12]. CVE-2025-43300, the Apple ImageIO vulnerability at the core of the chain studied here, is a CWE-787 instance.

CWE-863: Incorrect authorization [13] is categorically different. No bits are smashed and no memory is corrupted; the bug is a purely logical failure to enforce an access-control policy on a correctly authenticated principal. CVE-2025-55177, the WhatsApp linked-device synchronization bug, is a CWE-863 instance: the WhatsApp client trusts a synchronization message whose authorization scope was never fully verified, and the attacker uses this trust to coerce the victim device into processing attacker-controlled content. The relevance for this paper is that CWE-863 provides the *delivery* primitive that places the malicious image into the victim's media pipeline in the first place. Memory corruption alone is not enough for a zero-click exploit: the bug must be *reachable* without user interaction, and reachability is exactly what the authorization bug provides.

C. The iOS Media Processing Pipeline

On Apple platforms, almost every application that displays an image (Messages, Safari, Mail, WhatsApp, Signal) delegates decoding to a shared framework: ImageIO, together with its sister framework CoreGraphics. ImageIO is a C / Objective-C library responsible for parsing image container formats (JPEG, PNG, HEIF, DNG, GIF, TIFF, and several camera-specific RAW variants) and producing decoded pixel buffers. Because ImageIO is shared across the operating system, any parser bug inside it is reachable from every application that displays an image received from the network.

The critical observation is that ImageIO is invoked *automatically* the moment an incoming image enters the pipeline for thumbnail generation, *before* the user has seen, tapped, or acknowledged the message. This turns any parser bug in ImageIO into a zero-click attack primitive: the mere delivery of an image is sufficient to trigger code execution if the bug can be driven to memory corruption. In the WhatsApp–ImageIO chain, the end-to-end pipeline is:

1) WhatsApp receives a message via its linked-device synchronization protocol.
2) The message carries an image attachment.
3) To render a thumbnail, WhatsApp hands the raw bytes to CoreGraphics/ImageIO.
4) ImageIO dispatches to the parser for the declared format (DNG in this case).
5) The parser reads attacker-controlled header fields and triggers an out-of-bounds write.

None of these steps requires user interaction.

D. Modern Mitigations and Their Limits

Three decades of research have produced a layered defense against memory corruption [12], [14]. Three mitigations are most relevant for iOS and together define the landscape in which the 2025 chain had to operate.

Address Space Layout Randomization (ASLR) randomizes the base addresses of code and data regions at process start, so an attacker who wants to redirect control flow to a specific function must first *leak* an address from the victim process. ASLR does not prevent memory corruption; it raises the cost of turning corruption into useful code execution.

Stack canaries place a random value just before the saved return address on each stack frame. The function checks the canary before returning, and a mismatch terminates the process. Canaries protect the return address specifically, not arbitrary function pointers, not heap metadata, and not object vtables.

Application sandboxing runs each app inside a kernel-enforced container that restricts file-system, network, and IPC access. A successful RCE inside an app is therefore not immediately full device compromise: a separate sandbox escape is needed to reach kernel or cross-app territory.

These defenses raise the exploitation bar but leave structural gaps. ASLR can be bypassed with an information-leak primitive: often another parser bug in the same process [12]. Stack canaries do nothing for heap-resident function pointers, which are precisely what a heap overflow targets. Sandboxes reduce blast radius but do not prevent the initial compromise. The 2025 chain studied here operates *entirely within* these constraints: the parser bug is exploited inside a sandboxed process, and follow-on stages (privilege escalation, persistence) are layered on top.

E. Historical Precedent: FORCEDENTRY (2021)

In 2021, Citizen Lab and Google Project Zero documented *FORCEDENTRY* (CVE-2021-30860), a zero-click exploit deployed against Saudi activists and journalists via NSO Group's Pegasus spyware [8], [9]. The architecture is nearly identical to the 2025 chain studied in this paper:

- **Delivery:** the attacker sends an iMessage to the victim's phone number; no interaction is required.
- **Exploitation:** iMessage invokes CoreGraphics to render a malicious PDF disguised as a GIF. The PDF contains a JBIG2-compressed stream; a bug in the JBIG2 parser lets the attacker construct an arbitrary computation primitive inside the parser's memory, eventually achieving code execution.

- **Outcome:** full device compromise before the victim sees the message.

FORCEDENTRY proved that a memory-corruption bug in a shared image parser, combined with a delivery primitive that needs no user interaction, is sufficient for complete device takeover. Four years later, the 2025 WhatsApp–ImageIO chain reproduces the same architecture with different components: a different messenger (WhatsApp instead of iMessage), a different authorization bypass (linked-device sync confusion instead of direct delivery), a different parser (DNG in ImageIO instead of JBIG2 in CoreGraphics), and a different carrier format, but the same structural weakness and the same outcome. The persistence of this pattern across four years and across two independent messaging ecosystems is the central empirical observation of this paper, and motivates the technical analysis that follows.

III. Methodology

This paper is a technical case study, not an empirical measurement study. Our methodology is therefore organized around the evidence we use and the inferential framework we apply to it, rather than around an experimental protocol.

A. Case Selection Criteria

We selected the CVE-2025-55177 + CVE-2025-43300 chain as the primary case using four inclusion criteria:

1) **Real-world exploitation**, confirmed by inclusion in the CISA KEV catalog [3] and by contemporaneous reporting on civil-society victims.
2) **Public primary documentation** sufficient to reconstruct the chain without privileged access: vendor advisories, NVD entries, patch diffs, and at least one independent reverse-engineering report.
3) **C-class memory corruption at exploitation**, so that the case speaks directly to the research question.
4) **Zero-click end-to-end**, with no user action required between message receipt and code execution.

FORCEDENTRY (CVE-2021-30860) also satisfies all four criteria and is used as a historical comparison point (Section II-E) rather than as a second primary case. Isolated CVEs that require tap-to-open interaction, browser navigation, or installation of an attacker-controlled application were excluded, because their threat model differs qualitatively from the zero-click class studied here.

B. Analysis Framework

Each CVE is dissected along three axes. The *root cause* is the specific source-level condition (call site, allocation, missing check) that makes the bug possible. For CVE-2025-43300, this is the `SamplesPerPixel` vs. `SOF3` component-count mismatch in Apple's `CDNGLosslessJpegUnpacker`, reconstructed by Quarkslab from patch diffing of iOS 18.6.2 and macOS Sequoia 15.6.1 [10]. The *patch* is the vendor-supplied fix, treated as evidence of root cause: which check was added, which path was restricted, which validation was strengthened. Patch diffs are the most reliable external ground truth available to analysts without vendor access. *Position in the chain* locates the bug end-to-end: delivery primitive (CVE-2025-55177 grants reachability without interaction), exploitation primitive (CVE-2025-43300 turns reachability into remote code execution), or a glue step connecting the two.

The three-axis framework is applied symmetrically to both CVEs in Section IV and to FORCEDENTRY in Section II-E, enabling a like-for-like comparison across four years and two independent ecosystems.

C. Pedagogical Bridge: Phoenix `heap-two`

To make the exploitation mechanics legible to readers without prior heap-exploit experience, we conduct a full end-to-end walkthrough of the `heap-two` exercise from the Exploit Education Phoenix VM. The exercise is chosen for three reasons:

1) Its vulnerability class is CWE-787, the same class as CVE-2025-43300.
2) The primitive it grants (overwrite of a function pointer adjacent on the heap) is the closest textbook analogue of what the ImageIO bug ultimately yields.
3) It is small enough (50 lines of C) for the entire heap state to be inspected in `gdb` without losing the thread.

a) Setup: Installing and building `heap-two`.: The Exploit Education Phoenix VM is freely available from exploit.education. The `heap-two` source code resides in `/opt/phoenix/heap-two/heap-two.c`. We build the binary with:

```
gcc -o heap-two heap-two.c
```

By default, the educational binary is compiled without modern protections. We verify the mitigation state using `checksec`:

```
$ checksec --file=heap-two
```

Expected output: ASLR disabled, NX disabled, stack canary disabled, PIE disabled. This pedagogically useful configuration makes addresses deterministic and overflow mechanics observable without randomization obscuring the analysis.

b) Program structure and vulnerability.: The `heap-two` binary allocates two heap chunks and prompts for user input. The first chunk is a `char` buffer; the second is a struct containing a function pointer. A call to `strcpy` copies user input into the first buffer without bounds checking. Input longer than the allocated size causes overflow, corrupting the function pointer in the second chunk with an attacker-chosen value. Listing 2 shows the essential pattern.

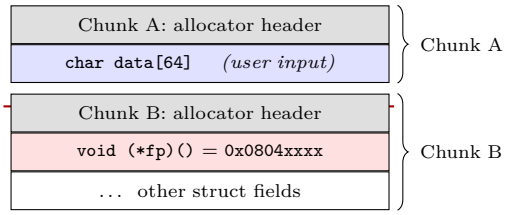

Fig. 1. Heap layout *before* overflow. Chunk A (64-byte user buffer) is adjacent to Chunk B (struct with function pointer). The dashed line marks the chunk boundary; a conforming write stays above it.

Listing 2. Illustrative heap-two structure. Two `malloc` calls produce adjacent chunks. `strcpy` performs no bounds check; input longer than `DATA_SZ` overwrites Chunk B's function pointer.

```
#define DATA_SZ 64

struct item    { char   data[DATA_SZ]; };
struct handler { void (*fp)(void);     };

struct item    *A = malloc(sizeof *A);
struct handler *B = malloc(sizeof *B);
B->fp = &safe_function;   /* legitimate default
     */

/* Unchecked: if strlen(input) > DATA_SZ, bytes
   overflow from A->data into B->fp
                    */
strcpy(A->data, user_input);

B->fp();  /* attacker controls this call target
     */
```

c) Execution workflow: `gdb` session.: We run the binary under `gdb` to observe heap layout and exploitation step-by-step:

1) Start the binary: `gdb ./heap-two`
2) Set breakpoint before vulnerable `strcpy`
3) Continue to breakpoint: `run`
4) Inspect heap state before overflow: `x/20x &chunk1`
5) Identify the function pointer location
6) Step through `strcpy`: `ni`
7) Re-examine memory after overflow
8) Observe the overwritten function pointer

d) Heap layout and the overflow.: Figure 1 shows the heap before overflow: two `malloc` calls produce chunks A (user buffer, 64 bytes) and B (struct with function pointer). When `strcpy` copies attacker input without bounds checking, bytes beyond chunk A's boundary overwrite chunk B's function pointer. Figure 2 shows the heap after overflow, with chunk B's pointer rewritten to the address of a `winner()` function.

e) Exploitation: crafting the payload.: The attacker constructs a payload filling chunk A with padding, then encodes the `winner()` function address in little-endian format to overwrite the pointer in chunk B. The program crashes at the attacker-chosen address: a successful exploit.

f) Connection to CVE-2025-43300.: The structural parallel is exact: both grant controllable heap write that overwrites an adjacent function pointer. Figures in

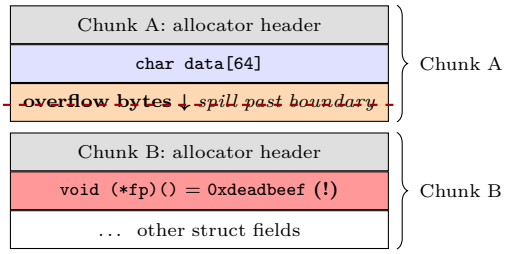

Fig. 2. Heap layout *after* overflow. Attacker input exceeds 64 bytes; excess bytes spill into Chunk B, overwriting the function pointer with an attacker-chosen value (`0xdeadbeef`). The red arrow shows the overflow path.

TABLE I
TOOLS AND PRIMARY SOURCES USED IN THIS STUDY.

Tool / source	Role
`gdb`	Heap and register inspection
`checksec`	Mitigation-state verification
Phoenix VM	Environment for `heap-two`
Hack The Box (Pwn track)	Supplementary heap/stack exploitation practice
Quarkslab report [10]	CVE-2025-43300 root cause
Apple, WhatsApp advisories	Canonical CVE descriptions
NVD entries	CWE class, CVSS, affected versions
CISA KEV [3]	Exploitation confirmation

Section IV present (i) a side-by-side diagram aligning `heap-two`'s layout with the ImageIO layout reconstructed from Quarkslab's analysis [10]; and (ii) screenshots of a `gdb` session showing the overwritten function pointer at the moment of control-flow transfer.

D. Tools and Primary Sources

All materials used in this study are publicly available; a summary is given in Table I. No proprietary data, no non-public code, and no new reverse engineering of Apple or WhatsApp binaries is produced or relied upon. The paper's analytical contribution is in synthesis and pedagogical framing, not primary discovery.

IV. Case Study: The 2025 Zero-Click Chain

In late August 2025, WhatsApp began notifying a set of users that their devices had been targeted in a "sophisticated cyber attack" combining a WhatsApp-specific authorization flaw (CVE-2025-55177) with a separately patched Apple ImageIO memory-corruption bug (CVE-2025-43300) [1], [2], [15]. The chain is the central object of study in this paper. Figure 3 overviews the end-to-end flow, which we dissect in three stages: delivery (Section IV-A), exploitation (Section IV-B), and the mapping from the textbook `heap-two` primitive to the real CVE (Section IV-C).

A. Stage 1 – Delivery: CVE-2025-55177

WhatsApp supports a multi-device model in which a primary account (a phone number registered on a phone)

Fig. 3. End-to-end flow of the 2025 zero-click chain. Red nodes mark the two patched CVEs. Only the victim's phone number is required; no action by the victim is needed for the chain to reach code execution.

can be paired with up to four additional endpoints: WhatsApp Web, Desktop, iPad, and so on. The feature is built on an extension of the Signal Protocol: the primary device holds the master identity key, and each linked device receives derived session material plus a stream of *synchronization messages* that replicate state changes (new chats, read receipts, media attachments) between devices [16].

The pairing handshake is the security-critical step. A legitimate pairing requires the primary device to scan a QR code displayed by the candidate linked device; the two devices then exchange public keys and the primary device records the new endpoint in its enrolled-device set. Synchronization messages are subsequently sent over the established secure channel and, by design, are authenticated against that enrolled set.

a) The bug.: Per WhatsApp's August 2025 security advisory [1], an "incomplete authorization of linked device synchronization messages" allowed a remote attacker to cause the victim's WhatsApp client to process sync content from a device whose enrollment had not been verified. Concretely, the client's handling of a subset of sync message types trusted an identifier embedded *in the message itself* to decide whether the sender was an authorized linked device, rather than consulting an independently maintained enrollment record. An attacker in possession of the victim's WhatsApp account identifier could therefore deliver crafted sync messages that the client would accept and process, including messages referencing attacker-controlled URLs for media attachments. The weakness is classified under CWE-863 (Incorrect Authorization) [13]; CVSS v3.1 scored 5.4 (Medium) on its own, without accounting for downstream impact when chained.

b) From authorization bypass to content processing.: The delivery path ends when the bypassed sync message instructs the victim's client to fetch and render an attacker-chosen image. WhatsApp, like most modern messengers, displays thumbnails for received images automatically in the chat-list preview; rendering invokes the system's shared image decoder (on iOS and macOS, CoreGraphics and ImageIO). The victim's client therefore downloads and parses the attacker's image *before* the victim has seen the message, producing the interaction-free attack surface on which Stage 2 operates.

c) Why this stage is zero-click.: Three properties together are what make Stage 1 zero-click in practice:

1) **No interaction to receive:** the sync message is delivered and processed by a background WhatsApp component, not by user-level code that requires the app to be foregrounded.
2) **No interaction to process attachments:** thumbnail generation runs at message arrival, not at message tap.
3) **No evidence before compromise:** the vulnerable code path executes before any visible UI change; the victim has no opportunity to decline.

d) Patch.: WhatsApp's fix, shipped in WhatsApp for iOS 2.25.21.73 and matching versions on Android, Desktop, and Business, tightened the authorization check on the affected sync message types so that only messages signed by a verifiably enrolled linked device are accepted [1]. After patching, the delivery primitive is closed: the exploitation primitive studied in Stage 2 remains reachable only through other code paths that feed attacker-controlled bytes to ImageIO (e.g., other messengers, browser download previews, Mail attachments).

B. Stage 2 – Exploitation: CVE-2025-43300

Once Stage 1 has forced the victim's device to fetch and render an attacker-chosen image, the second CVE takes over. CVE-2025-43300 is a heap out-of-bounds write (CWE-787) in Apple's ImageIO framework, specifically in the handling of lossless-JPEG streams embedded inside DNG (Digital Negative) raw image containers. It was patched by Apple on 20 August 2025 in iOS 18.6.2 and macOS Sequoia 15.6.1 with the advisory note that "Apple is aware of a report that this issue may have been exploited in an extremely sophisticated attack against specific targeted individuals" [2].

a) Container format: DNG.: DNG is a TIFF-based container defined by Adobe for raw camera images. A DNG file consists of one or more Image File Directories (IFDs), each carrying tagged metadata describing the stored pixel data. Two tags are directly relevant:

- `SamplesPerPixel` (tag 277): the number of samples per pixel, typically 1 for single-channel grayscale or 3 for RGB.
- `Compression` (tag 259): when set to 7 (lossless JPEG), the pixel payload is a lossless-JPEG-compressed stream embedded inside the TIFF and referenced through `StripOffsets`/`TileOffsets`.

Inside the lossless-JPEG stream, the component count is given *independently*, by the `Nf` field of the SOF3 (Start of Frame 3, lossless JPEG) marker. In a well-formed DNG, `Nf` and `SamplesPerPixel` must agree: the container-level count of samples per pixel and the stream-level count of JPEG components describe the same quantity.

b) The bug.: The ImageIO decoder for this configuration is implemented in `CDNGLosslessJpegUnpacker` (private to ImageIO; symbol names here follow Quarkslab's reconstruction [10]). The decoder allocates its output buffer using `SamplesPerPixel` taken from the TIFF IFD, then drives the decoding loop using `Nf` taken from the SOF3 marker of the embedded lossless JPEG. When an attacker crafts a DNG in which `SamplesPerPixel = 1` but `SOF3.Nf = 3`, the decoder allocates a buffer sized for one component and writes three components' worth of decoded data into it, overflowing by a factor of three [10].

The overflow is useful because it is:

- **Controllable in size:** the attacker chooses both `SamplesPerPixel` and `Nf`, so the overflow length is set by simple arithmetic on attacker-chosen inputs.
- **Controllable in content:** the overflowing bytes are derived from the lossless-JPEG payload, which is attacker-chosen end to end.
- **Heap-adjacent:** the target allocation lives on the heap used by ImageIO for pixel buffers; adjacent objects on that heap include decoder state, buffer metadata, and function pointers that are dereferenced by subsequent decoding code.

Quarkslab summarizes the pattern as "two bytes that make size matter" [10]: the disagreement between a single-byte count in the TIFF header and a single-byte count in the SOF3 marker is the entire bug.

c) From overflow to code execution.: The gap between "heap OOB write" and "code execution on an iPhone" is large in general but has a well-understood shape. First, shape the heap (via prior allocations performed by the parser itself) so that the object immediately after the victim buffer is one whose type contains a function pointer that the decoder will later call. Second, overflow precisely enough to overwrite that pointer without corrupting intervening metadata. Third, when the decoder later invokes the overwritten pointer, control transfers to the attacker-chosen address.

On iOS, these steps are further constrained by ASLR, by pointer authentication codes (PAC) on ARMv8.3+ Apple silicon that sign pointers with a process secret, and by the app sandbox which limits the power of a code-execution primitive inside ImageIO. Full Stage 2 exploitation in the wild therefore requires additional primitives (information leaks, PAC bypasses, sandbox escapes) that are not part of CVE-2025-43300 itself but are necessary in practice. For the purpose of this paper, the essential observation is that the *structural* primitive granted by CVE-2025-43300 is the same one granted by the textbook `heap-two` exercise: overflow a heap buffer into adjacent attacker-reachable state, then hijack control flow when the program dereferences the corrupted state.

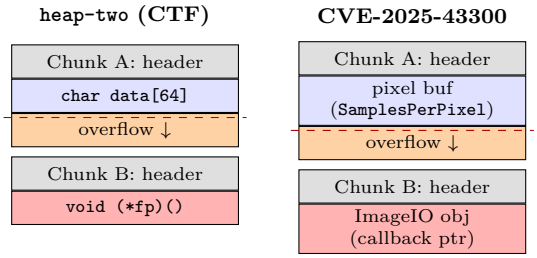

Fig. 4. Side-by-side heap layout. Left: `heap-two`: chunk A holds an attacker-copied string, chunk B holds a function pointer `fp`; `strcpy` overflows A into B. Right: CVE-2025-43300, where chunk A is the pixel buffer sized by `SamplesPerPixel`, chunk B is an internal ImageIO object with a dereferenced callback; the SOF3-driven decode loop overflows A into B.

d) Patch.: Apple's fix, inferred from diffing iOS 18.6.1 against iOS 18.6.2 per Quarkslab's analysis [10], adds a consistency check ensuring that `SamplesPerPixel` and the embedded `SOF3.Nf` agree before the decoding loop begins. Inputs for which the two quantities disagree are rejected at parse time; no allocation/decoding mismatch is reachable on the patched code path. The patched binary adds on the order of a few bytes of validation to prevent an overflow whose severity is measured in complete device compromise, an asymmetry common to parser bugs of this shape.

C. From CTF to CVE

At the level of exploit primitive, `heap-two` and CVE-2025-43300 are the same bug. Both grant an attacker a controllable heap write that extends past the allocated bounds of an attacker-sized buffer into attacker-reachable adjacent state. Both terminate in a function-pointer overwrite that is later dereferenced by the vulnerable program. The difference between the 20-line Phoenix binary and a production zero-click chain is engineering, not mechanism.

a) The mechanism, aligned.: Figure 4 aligns the two heap layouts. In `heap-two`, `malloc` is called twice; the first chunk is the user-controllable buffer and the second contains a function pointer `fp` that the program later calls. The call to `strcpy` copies attacker-controlled data into the first chunk without respecting its allocated size, and bytes past the end of that chunk fall exactly onto `fp` in the second chunk. Overwriting `fp` with the address of the target function `winner()` redirects execution. In CVE-2025-43300, the first chunk is the pixel buffer sized by `SamplesPerPixel`; the second chunk is, depending on heap shape, an internal ImageIO object whose virtual-method pointer or callback is subsequently dereferenced during image processing. The rest is the same.

b) What scales from textbook to production.: Productionizing the primitive focuses on two engineering problems. First, the textbook exercise runs with predictable

[Figure placeholder: `gdb` session on `heap-two`]

Fig. 5. `gdb` session at the moment of the overwritten function pointer in `heap-two`. The target chunk address and the overwriting bytes are visible in the heap dump (left); the corrupted function pointer is visible in the register state after the `call` instruction (right).

addresses because ASLR is disabled in the educational binary; a real exploit must first acquire an information leak, often by exploiting a second, milder bug in the same process, to locate its target. Second, the textbook exercise enjoys a deterministic heap layout because `malloc` is called twice in the attacker's order; the real exploit must shape a live, concurrent allocator (`libmalloc` on iOS) so that the attacker's target object reliably lands immediately after the overflowable buffer. The standard technique is *heap grooming*: forcing the vulnerable parser to allocate a series of predictable-size buffers that reshape free-list structure, then triggering the vulnerable allocation at a point where the target object is the next chunk the allocator will return.

c) What does not scale.: The bug class itself (out-of-bounds write on the heap with attacker control over size and content) is identical across cases. This is the core observation driving the research question: the fundamental memory-corruption mechanism that a student can learn in an afternoon on an educational VM is the same one that compromised approximately 200 mobile devices in mid-2025. Mitigations (ASLR, heap randomization, PAC, sandboxing) raise the cost of Stage 2 exploitation but do not alter the structural vulnerability. Section V examines why this mitigation model has structural limits.

V. Discussion

The 2025 WhatsApp–ImageIO chain and its 2021 FORCEDENTRY predecessor demonstrate a consistent architectural pattern: a delivery primitive (messaging protocol bypass) composes with an exploitation primitive (image parser memory corruption) to achieve zero-click code execution. This section examines why the current mitigation stack did not prevent the 2025 chain, what lies beyond the initial RCE, what structural alternatives exist, and the limitations of this study.

A. Why the Mitigation Stack Did Not Prevent This Chain

Section II-D introduced three mitigations deployed on iOS: ASLR, stack canaries, and application sandboxing. None of them target the root cause of CVE-2025-43300.

ASLR randomizes memory layout but does not prevent the out-of-bounds write itself. The write occurs regardless of where the buffer is placed; ASLR only makes it harder to know where to redirect control flow afterward. In practice, a second, milder bug in the same process (an information leak) suffices to defeat ASLR by revealing a code or data address, after which the attacker can construct a valid target for the hijacked pointer. Information leaks in image parsers are common: the same complexity that produces out-of-bounds writes routinely produces out-of-bounds reads [12].

Stack canaries are entirely irrelevant: the overflow occurs on the heap, not the stack. Canaries protect saved return addresses; they do not monitor heap metadata or function pointers stored in heap-allocated objects.

Application sandboxing limits what the attacker can do after RCE but does not prevent the RCE itself. The sandbox around WhatsApp restricts file-system and IPC access, but the attacker's initial foothold (code execution inside the process) is already sufficient to exfiltrate data accessible to that process (contacts, messages, media, location) and to attempt a sandbox escape via a separate vulnerability.

Pointer Authentication Codes (PAC), deployed on ARMv8.3+ Apple silicon [17], sign pointers with a process-specific secret so that corrupted pointers fail verification before use. PAC raises the cost of control-flow hijacking significantly but has been bypassed in academic and in-the-wild settings, typically by reusing legitimately signed pointers or by targeting code paths that do not enforce PAC on every indirect call.

The pattern is consistent: each mitigation addresses a *symptom* (where code lives, whether the return address is intact, what permissions the process has, whether a pointer is signed) rather than the *root cause* (the program wrote past the end of a buffer because no bounds check existed). So long as the parser is written in a language that permits unchecked memory access, the vulnerability class persists and mitigations can only raise the cost of exploitation, not eliminate it.

B. Beyond RCE: Sandbox Escape and Persistence

A zero-click RCE inside a sandboxed process is not, by itself, full device compromise. The reported FORCEDENTRY and 2025 chains both achieved persistent, cross-application surveillance (call recording, message exfiltration, camera/microphone access), implying that additional stages followed the initial RCE: at minimum, a sandbox escape and a privilege-escalation step. These stages are outside the scope of this paper: they involve kernel vulnerabilities and system-daemon bugs unrelated to the two CVEs studied, but their existence must be acknowledged: the image-parser RCE is the entry point, not the endpoint, of the full attack.

C. Structural Alternatives

If mitigations cannot close the vulnerability class, what can? Two categories of structural change address the root cause.

Memory-safe languages (Rust, Swift, Go, Java) enforce bounds checking, type safety, and lifetime management at compile time or runtime, making out-of-bounds writes impossible in safe code by construction [7], [18]. Rewriting media parsers in a memory-safe language would eliminate CWE-787 from the attack surface entirely. The industry is already moving in this direction. Apple has begun shipping Swift-based replacements for some system components. Google's Android team has documented that the share of memory-safety vulnerabilities in new Android code fell to near zero in modules rewritten in Rust [7]. Most significantly, the Linux kernel (the canonical example of a safety-critical C codebase) merged initial Rust infrastructure in Linux 6.1 (December 2022) and shipped its first production Rust-written driver (a network PHY driver for the Asix ax88796b chipset) in Linux 6.8 (March 2024); Rust was declared a permanent, non-experimental part of the kernel at the 2025 Kernel Maintainer Summit in Tokyo [19]. At the distribution level, Ubuntu 26.04 LTS (April 2026) became the first major long-term-support release to ship *sudo-rs*, a memory-safe Rust reimplementation of `sudo`, as the *default* privilege-escalation binary [20], replacing a C codebase that has been a high-value target for local privilege escalation for decades. The challenge for media frameworks is scale: ImageIO is a large, performance-critical C/Objective-C codebase with decades of accumulated optimisation, and incremental rewriting is slow, but sudo-rs and the Linux Rust drivers demonstrate that production-grade, security-critical C code can be incrementally replaced.

Hardware-enforced memory safety, specifically ARM Memory Tagging Extension (MTE) on ARMv8.5+, assigns a 4-bit tag to every 16-byte memory granule and checks the tag on every access. A heap overflow that crosses a granule boundary with the wrong tag traps immediately, preventing the corruption from reaching adjacent objects. MTE is deployed on some Android devices but is not yet enabled on iOS. If deployed, MTE would make the CVE-2025-43300 overflow detectable at the hardware level regardless of the implementation language.

D. Limitations

Three limitations must be acknowledged. First, our analysis of CVE-2025-43300 is second-hand: we rely on Quarkslab's reconstruction from patch diffing [10], not on independent reverse engineering. Certain exploitation details (heap-spraying strategy, PAC bypass technique) are inferred, not observed. Second, the Phoenix `heap-two` exercise deliberately omits real-world complications (ASLR, PAC, heap randomization, concurrent allocation); it illustrates the core mechanic, not the full exploitation engineering. Third, we compare only two chains (FORCEDENTRY and the 2025 chain); broader generalization of the "same pattern, different parser" thesis requires analysis of additional cases as they are disclosed.

VI. Conclusion

This paper analyzed the 2025 WhatsApp–ImageIO zero-click exploit chain (CVE-2025-55177 + CVE-2025-43300), which compromised the mobile devices of fewer than 200 targeted individuals without any user interaction. Through a three-axis framework (root cause, patch, chain position), we reconstructed how a WhatsApp linked-device authorization bypass (CWE-863) delivered a crafted DNG image to Apple's ImageIO framework, where a two-byte inconsistency between TIFF metadata and an embedded JPEG marker triggered a heap out-of-bounds write (CWE-787) leading to remote code execution.

Three findings emerge. First, the fundamental exploitation primitive (a heap buffer overflowed into an adjacent function pointer) is identical to the one demonstrated in the educational Phoenix `heap-two` exercise. The difference between a textbook CTF exercise and a production zero-click chain is engineering (heap grooming, ASLR bypass, PAC bypass), not mechanism. Second, the 2025 chain reproduces the architectural pattern of FORCEDENTRY (2021) with different components (a different messenger, a different parser, a different image format) but the same structural vulnerability. The recurrence across four years and two independent ecosystems suggests that the pattern is inherent to the current architecture of media processing on mobile platforms, not an isolated accident. Third, the deployed mitigation stack (ASLR, stack canaries, sandboxing, PAC) addresses symptoms rather than the root cause: none of these defenses prevent the out-of-bounds write itself.

We recommend two structural interventions. The near-term priority is incremental rewriting of high-exposure C/C++ parsers (ImageIO, CoreGraphics, and their Android equivalents) in memory-safe languages such as Rust or Swift. The longer-term intervention is deployment of hardware-enforced memory tagging (ARM MTE) on iOS, which would make heap overflows detectable at the granule level regardless of the implementation language.

Future work should track whether the "delivery + parser" pattern continues to appear in newly disclosed zero-click chains, and should extend the Phoenix-to-CVE pedagogical bridge to additional vulnerability classes (use-after-free, type confusion) as suitable educational exercises become available.

References

[1] WhatsApp / Meta, "WhatsApp Security Advisory — CVE-2025-55177," Official advisory, August 2025. [Online]. Available: https://www.whatsapp.com/security/

[2] Apple Inc., "Apple Security Updates — CVE-2025-43300," Official security advisory, August 2025. [Online]. Available: https://support.apple.com/en-us/HT201222

[3] CISA (Cybersecurity and Infrastructure Security Agency), "Known Exploited Vulnerabilities Catalog," Online catalog, 2025. [Online]. Available: https://www.cisa.gov/known-exploited-vulnerabilities-catalog

[4] ResearchGate, “Technical Research Report: Advanced iPhone Zero-Click Exploits and Spyware (2019–2025),” Research report, 2025. [Online]. Available: https://www.researchgate.net/
[5] A. One, “Smashing the Stack for Fun and Profit,” *Phrack Magazine*, vol. 7, no. 49, November 1996. [Online]. Available: http://phrack.org/issues/49/14.html
[6] MITRE Corporation, “CWE-787: Out-of-bounds Write,” Online database, 2024. [Online]. Available: https://cwe.mitre.org/data/definitions/787.html
[7] CISA, “The Urgent Need for Memory Safety in Software Products,” Policy document, November 2023. [Online]. Available: https://www.cisa.gov/news-events/alerts/2023/11/14/urgent-need-memory-safety-software-products
[8] Citizen Lab, “FORCEDENTRY: NSO Group iMessage Zero-Click Exploit Captured in the Wild,” Research report, September 2021. [Online]. Available: https://citizenlab.ca/2021/09/forcedentry-nso-group-imessage-zero-click-exploit-captured-in-the-wild/
[9] Google Project Zero, “A Deep Dive into an NSO Zero-Click iMessage Exploit: Remote Code Execution,” Research blog, December 2021. [Online]. Available: https://googleprojectzero.blogspot.com/2021/12/a-deep-dive-into-nso-zero-click.html
[10] Quarkslab, “Reverse Engineering of Apple’s iOS 0-click CVE-2025-43300: Two Bytes That Make Size Matter,” Research blog, August 2025. [Online]. Available: https://blog.quarkslab.com/
[11] Anonymous, “Once upon a free(),” *Phrack Magazine*, vol. 11, no. 57, August 2001. [Online]. Available: https://phrack.org/issues/57/9
[12] L. Szekeres, M. Payer, M. Dell’Amico, and S. Devadas, “SoK: Eternal War in Memory,” *IEEE Symposium on Security and Privacy (S&P)*, pp. 48–62, 2013.
[13] MITRE Corporation, “CWE-863: Incorrect Authorization,” Online database, 2024. [Online]. Available: https://cwe.mitre.org/data/definitions/863.html
[14] V. van der Veen, L. Cavallaro, G. Herbert, C. Kruegel, and G. Vigna, “A Survey of Prevention/Mitigation Techniques against Memory Corruption Attacks,” *ACM Computing Surveys*, vol. 48, no. 3, pp. 1–35, 2016.
[15] Lookout Threat Intelligence, “CVE-2025-55177: WhatsApp Linked Device Authorization Bypass,” Threat report, August 2025. [Online]. Available: https://www.lookout.com/
[16] WhatsApp, “WhatsApp Security and Privacy Overview,” Technical whitepaper, 2023. [Online]. Available: https://www.whatsapp.com/security/
[17] ARM Limited, “Pointer Authentication on ARMv8.3,” Architecture documentation, 2019. [Online]. Available: https://developer.arm.com/documentation/
[18] M. Kaminski, “Memory Safety in C/C++ at Scale,” *USENIX Login*, vol. 44, no. 4, 2019.
[19] LWN.net / Phoronix, “Rust Drivers Land in Linux 6.8; Rust Declared Permanent at 2025 Kernel Maintainer Summit,” Online news, 2025. [Online]. Available: https://www.phoronix.com/news/Linux-6.8-Rust-PHY-Driver
[20] Canonical / Prossimo (ISRG), “Ubuntu 26.04 LTS Ships sudo-rs as the Default Privilege-Escalation Tool,” Online announcement, April 2026. [Online]. Available: https://www.memorysafety.org/blog/sudo-rs-headed-to-ubuntu/