

# Lest We Forget: Cold-Boot Attacks on Scrambled DDR3 Memory

Bauer, Gruhn, Freiling

Universität Erlangen-Nürnberg

March 30th 2016

# Preface

- ▶ “Lest we Remember”: Halderman et al. 2009, alludes to Isaac Asimov’s short story; protagonist achieves perfect memory by use of a drug
- ▶ “Lest we Forget”: alludes to Rudyard Kipling’s poem “Recessional”; warns not to forget quickly
- ▶ The point we’re trying to make: cold boot attacks are still working even with modern memory technologies

# Forensic Memory Acquisition

- ▶ RAM contains lots of evidence of forensic interest (e.g. TLS session keys, FDE keys, evidence of resident rootkits, etc)
- ▶ A snapshot of RAM can be captured either in software (on a running system) or in hardware (on the same or a *different* system)
- ▶ Both approaches have their distinctive use-cases in which they're applicable, both have up- and downsides

# Forensic Memory Acquisition

- ▶ Cold boot attack: Hard reset of the system and booting into a minimal, memory-dumping OS or transplanting the memory IC into a different PC
- ▶ Gruhn/Müller 2013, On the Practicability of Cold Boot Attacks: "However, we also point out that we could not reproduce cold boot attacks against modern DDR3 chips."

# What is RAM?

- ▶ RAM refers to memory which
  - ▶ has low latency (typ. 5-20 ns range)
  - ▶ provides great bandwidth (typ. 10-50 GB/s)
  - ▶ is usually volatile
- ▶ There are a **lot** of different technologies
- ▶ We focus on SDRAM that is widely used in computers today:  
DDR3
- ▶ DDR3 uses capacitor-based bit storage

# DDR2 vs. DDR3

- ▶ DDR2 and DDR3 are very similar technologies
- ▶ Roughly speaking, DDR3 provides greater throughput at the cost of higher latency
- ▶ This is achieved by doubling the minimum burst length of a memory transaction
- ▶ With increasing speed, managing the required charging/discharging of cells becomes increasingly difficult

# DDR3 approach

- ▶ The MCH (component in the CPU that talks directly to RAM) was therefore improved by Intel starting with DDR3 generations
- ▶ Basic idea: XOR the datastream with a PRNG pattern (that's called scrambling or *whitening*)
- ▶ → storage bitstream in which statistically half of the bits are set and half are cleared
- ▶ i.e. always the *average* case, mitigating the  $\frac{d}{dt}$  peaks
- ▶ Hamming-weight of data in memory is statistically zero-sum (i.e. free of bit bias)

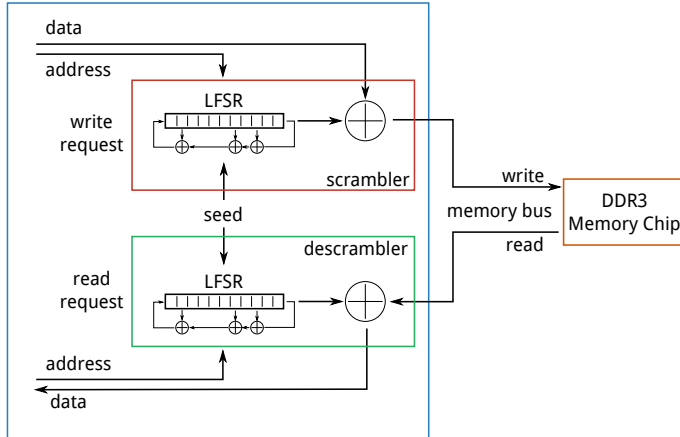
# The R in RAM

- ▶ To be able to still *randomly* access memory, the scrambler unit needs to be able to *seek* to parts of the PRNG stream
- ▶ Intel approach: Use LFSRs, parametrize the LFSRs with a *seed* value that easily allows jumping to an arbitrary location
- ▶ → No performance penalty by scrambler
- ▶ Rough explanation in the Intel patent on that subject



# Schematically

Memory Controller Hub (within CPU)



# Implications

- ▶ If you use a DDR3 system to capture RAM content, you'll only ever see scrambled images
- ▶ In fact, those images will have been scrambled *twice* (once by the scrambler and then again by the descrambler)
- ▶ One of our approaches therefore looked at *differential images*

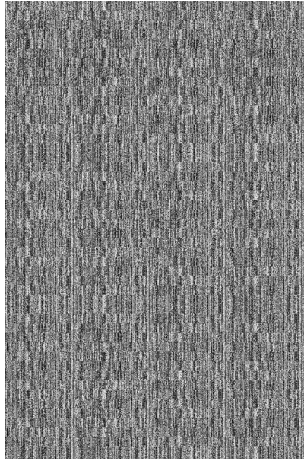
# Towards descrambling

- ▶ The Intel patent is vague and encompasses a lot of different technologies
- ▶ Which one is actually used is not described (i.e. how many parallel LFSRs, which bitlength, which bit order, etc.)
- ▶ Docs are unavailable (to us), only few lines of reverse-engineered CoreBoot code available

# Reprogramming the scrambler

- ▶ The BIOS is usually stored on Flash, (approx. 100ns to read a single byte at 80 MHz)
- ▶ Therefore, RAM initialization happens *very early* during boot
- ▶ In fact, it's one of the first things the BIOS does so it can relocate itself from Flash to DRAM
- ▶ We have good reason to believe that reprogramming the scrambler with an active RAM channel is disallowed by hardware (as it should be)
- ▶ This makes probing within the system and hacking code difficult

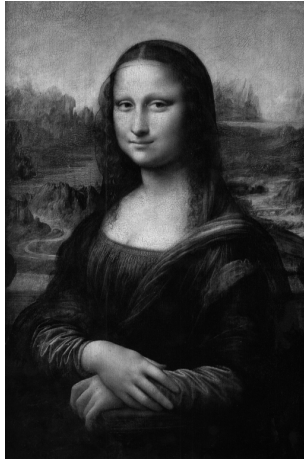
# Looking at dumped ground-state memory



# Ground state problem

- ▶ The ground pattern of a memory chip is undefined
- ▶ It tends to correlate, however, with its physical construction (i.e. if cells are biased against  $V_{cc}$  or GND)
- ▶ It is therefore difficult to extract the stream cipher key crumbs (because the plaintext varies)

# Placing Mona Lisa in that spot

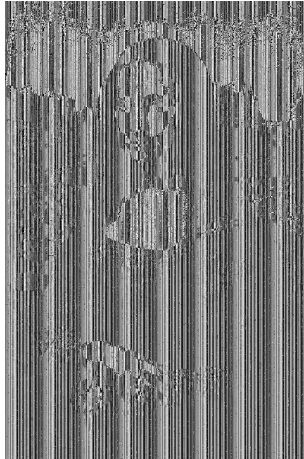


# Use freeze spray





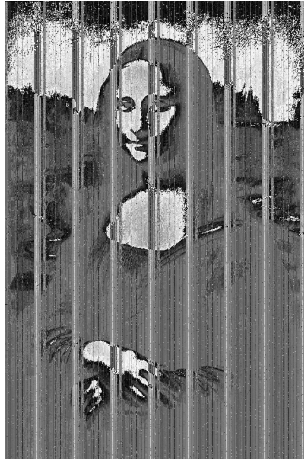
# First cold-boot Lisa memdump (-30°C)



# Mona Lisa memdump

- ▶ It's clearly visible that the information survived the reboot (i.e. no clearing of the memory was performed during initialization)
- ▶ And we additionally know the plaintext
- ▶ But if we didn't, we could first try to descramble it with a related-key approach

# Using related-key descrambling



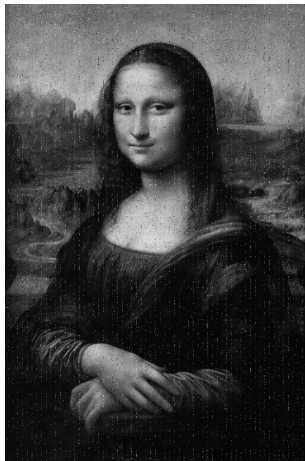
# Systematic approach

- ▶ XORing the memdump with the original image gives us an approximation of the PRNG stream
- ▶ We partition this PRNG stream in chunks of different length (spoiler: 64 bytes)
- ▶ And try to group them together in groups which have many bits matching (because of acquisition errors)
- ▶ Then do majority voting on the key bits (most in agreement likely to be correct)

# Systematic approach

- ▶ This allows us to extract the two 64-bytes keys (one for each memory channel)
- ▶ With deinterleaving (which we also describe in the paper) the image can now successfully be descrambled
- ▶ By utilizing the LFSR construction and congruencies which we noticed within the keystream, we reduce that known plaintext from 128 bytes to just 50 bytes

# Final result



# Overt observations

- ▶ The obvious observation is that we can descramble transplanted memory using the described method and with 50 bytes of known plaintext
- ▶ A side note should be that capturing DDR3 snapshots is much more difficult than with DDR2 memory (much higher bit decay)
- ▶ It was more difficult than we imagined to get accurate results

# Covert observation

- ▶ Thinking out loud: Intel implemented this to mitigate the detrimental effects of  $\frac{d}{dt}$  peaks
- ▶ If you know the scrambler stream, however, it is still perfectly possible to force such peaks
- ▶ Fill a buffer with the keystream, invert it repeatedly.
- ▶ Could this possibly lead to memory corruption attacks like rowhammer introduced? Not sure, but surely tempting.



# Are there any...

# ...questions?