CS 489 / 698: Software and Systems Security

Module 2: Program Security (Attacks)
weird machine

Meng Xu *(University of Waterloo)*

Spring 2023

Outline

Based on paper

Weird Machines, Exploitability, and Provable Unexploitability

By *Thomas Dullien* published in 2017 when he was in Google Project Zero.

Why this paper?

It attempts to formalize a concept that has been intuitively known for quite a while in the community of security practitioners, i.e., both by the hackers and the researchers...

... and that concept is called "exploit".

What is an exploit?

- Magic
- Access (mostly unauthorized)
- Controls the instruction pointer (e.g., EIP/RIP register)
- A program does something it is not supposed to do
- I can recognize it when I see it

They are not technically wrong, but are clearly ill-defined for academic research purposes.

Why do we bother to define it?

We need to make justifications in the real-world that depends on the concept of "exploits":

- Mitigation strategies
  - e.g., difficulty of exploitation vs performance
  - e.g., difficulty of exploitation vs programmability
  - e.g., difficulty of exploitation vs complexity
- Exploitability of software/hardware defects
  - e.g., does the Rowhammer bug makes a big security problem?
  - e.g., can the Spectre bug be used to launch general attacks?
  - e.g., if yes, how?

The MitiGator

Raising the bar on exploitation until no more exploits can be seen

Learn principles, not examples

An important message conveyed by this paper (which is also a message I want to share with you), is that exploitation IS NOT a "bag of tricks".

In security courses (including this one), we teaches

- Stack smashing, buffer overflows, heap exploitations
- SQL injection, XSS, etc
- ASLR, CFI, sandboxing, etc.

It is important to remember that there is a more fundamental principle behind these examples — *exploitation is all about entering and programming a weird machine*.

Outline

Behind an "exploit"

By just saying that "I exploited something", you are conveying at least two messages:

- There exists some software running on top of some hardware
- There are "defects" in either the software or hardware (or both).

What is software?

A software is an emulator for a finite-state machine (FSM) we would like to have but we don't.

Instead, we only have a general-purpose CPU which is designed to model a huge spectrum of FSMs.

Hence, the reason we develop software is to confine the CPU to follow and only follow the FSM we intend to have.

The intended finite state machine (IFSM)

The state machine we want to have is called the “intended finite-state machine” (IFSM).

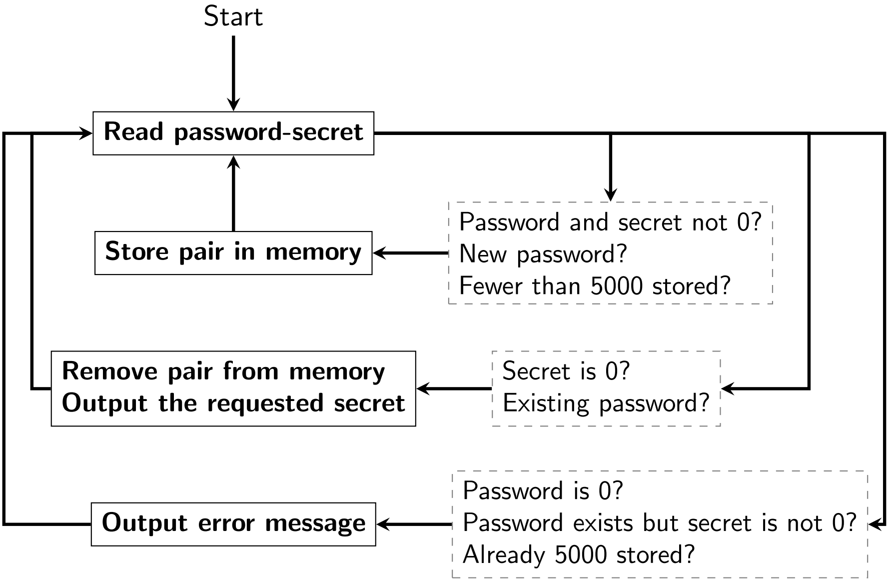
- It is usually not explicitly specified
- It is “perfect” by design — fully implements our intents
- It cannot, by definition, have security problems.

A concrete example: a secret-keeping machine

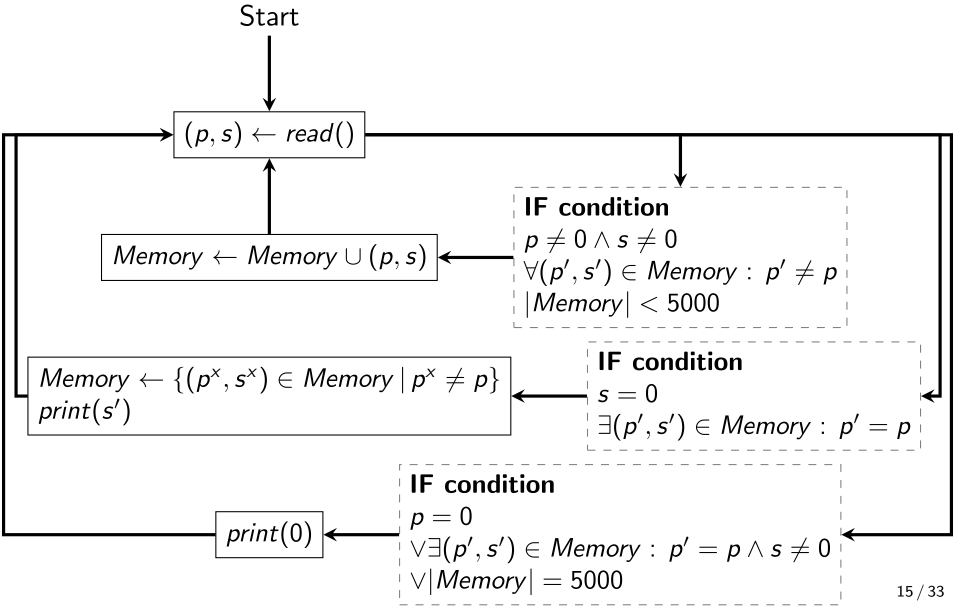
The machine has the following functionalities:

- Reads a password / secret (p, s) from a user and remembers it.
  - NOTE: neither p nor s can be 0 (0 is reserved as an error code)
- Given a password (p) that exists in the memory, the machine returns a previously-stored secret (s) and forget both.
- The machine will not need to store more than 5000 such pairs.

IFSM diagram



IFSM diagram



IFSM formalization

The set of all *Memory*, denoted as $\mathcal{M}$, can be formally defined as

$$\mathcal{M} = \left\{ \begin{array}{l} \emptyset \\ \{(p_1, s_1)\} \\ \ldots \\ \{(p_1, s_1), \ldots, (p_{5000}, s_{5000})\} \end{array} \middle| \begin{array}{l} p_i, s_i \in \{0,1\}^{32} - \{0\} \\ p_i \neq p_j \\ \\ \\ \end{array} \right\}$$

FSM quick recap

An FSM can be defined by a 7-tuple: $(Q, i, F, \Sigma, \Delta, \delta, \sigma)$

- Q: Set of states
- i: The initial state
- F: The set of final states
- Σ: The input alphabet
- Δ: The output alphabet
- δ: State transition function $\delta : Q \times \Sigma \rightarrow Q$
- σ: Output mapping function $\sigma : Q \times \Sigma \rightarrow \Delta$

IFSM formalization — what we intend to have

The IFSM of our secret-keeping program can be defined as:

- Q: $\{A_M, M \in \mathcal{M}\}$
- i: $A_\emptyset$
- F: $\emptyset$
- Σ: $\{(p, s) \mid p, s \in \{0, 1\}^{32}\}$
- Δ: $\{0, 1\}^{32}$
- δ: $A_M \times (p, s) \to A_M \mid A_{M \cup (p,s)} \mid A_{M-(p,s)}$
- σ: $A_M \times (p, s) \to s' \mid 0$

What we actually have: a realistic CPU

The Cook-and-Reckhow RAM machine

- 2^{16} memory cells each holding a 32-bit value
- 7 CPU registers (r_0 to r_6)
- A small set of instructions
  - Constant: LOAD(C, r_d)
  - Register operations: ADD(r_{s1}, r_{s2}, r_d)
  - Register operations: SUB(r_{s1}, r_{s2}, r_d)
  - Memory read: ICOPY(r_p, r_d)
  - Memory write: DCOPY(r_d, r_s)
  - Control flow: JNZ/JZ(r, l_z)
  - Environment IO: READ(r_d)
  - Environment IO: PRINT(r_s)
- Harvard architecture (program is provided and external to RAM)

CPU FSM formalization — what we actually have

The FSM of a general-purpose CPU can be defined as:

- Q: $(q_1, ..., q_{2^{16}}) \times (r_0, ..., r_6) \times p_i$ where $q_i, r_i \in \{0,1\}^{32}$, $p_i \in P$
- i: $q_i = 0, r_i = 0, p_i = P_0$
- F: $\emptyset$
- Σ: CPU Instruction Set $\{I\}$
- Δ: $\{0,1\}^{32}$
- δ: $Q \times I \rightarrow Q'$
- σ: $Q \times I \rightarrow (e \in \Delta)$

From spec to execution: a series of refinement

We want to translate our IFSM S_{spec} into our CPU FSM $S_{execution}$.

It is actually a multi-stage process, involving (non-exhaustively)

$$S_{spec} \sqsupseteq S_{language} \sqsupseteq S_{machine} \sqsupseteq S_{execution}$$

- $S_{spec} \not\sqsupseteq S_{language}$: software bug, blame the developer
- $S_{language} \not\sqsupseteq S_{machine}$: compiler bug, blame the compiler
- $S_{machine} \not\sqsupseteq S_{execution}$: hardware bug, blame the machine

Outline

Bug $\implies$ exploits?

Does having a bug in the refinement chain always implies a security issue (a.k.a., an exploit)?

What is security?

Security are properties of the IFSM that we want to hold in the presence of an adversary with a specific attack model.

Security of our secret-keeper

Informally, we want to ensure that anyone who interact with our program need to know (or guess) the right password in order to obtain the stored secret.

Put in a different way, the best way to attack our program to extract some secret is to guess the password.

Security of our secret-keeper

Formally, we want the security property to hold at our IFSM:

$$Pr[s \in O_{IFSM}] \leq \frac{|I_{attempt}|}{2^{32}}$$

As well as at the final execution stage, after the refinement chain

$$Pr[s \in O_{execution}] \leq \frac{|I_{attempt}|}{2^{32}}$$

Even in the presence of an attacker with the assumed power of performing single chosen bit-flip.

The security property depends on the implementation

- Naive implementation: Simulate the *Memory* set as a flat linear array with sequential scanning
- Clever implementation: Simulate the *Memory* set with two singly-linked lists.

Conclusion: the clever implementation is actually vulnerable.

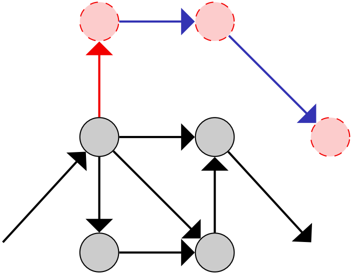
An attack on the clever implementation

1. Attacker sends (p_0, s_0), (p_1, s_1), (p_2, s_2)
2. Victim sends (p_d, s_d)
3. Attacker sends $(p_2, 0)$, $(p_1, 0)$, (p_3, s_3), (p_4, s_4)
4. Attacker gets to corrupt a single bit: flip the least significant bit for memory cell content at `b'0101` (i.e., cell `0x5`)
5. Attacker sends $(s_4, 0)$
6. Attacker sends $(12, 0)$ and obtains s_d

The naive implementation is secure

Please refer to the paper for the details of the proof.

Programming the weird machine

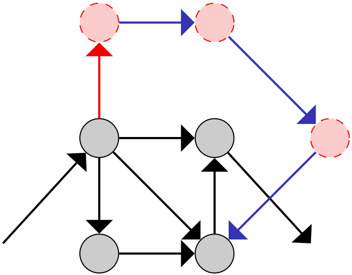


An emergent insruction set

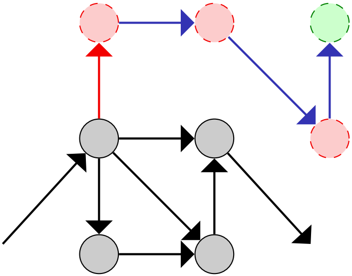
This weird machine creates an emergent instruction set that is constrained by:

- The IFSM
- The program that is refined from the IFSM
- The CPU FSM

Outcomes of weird machine programming



Reverted back to the IFSM



Reached the target state

$\langle$ **End** $\rangle$