CSci 5271
Introduction to Computer Security
Day 5: Low-level defenses and counterattacks
Stephen McCamant
University of Minnesota, Computer Science & Engineering

Outline
Return address protections
Announcements intermission
ASLR and counterattacks
W⊕X (DEP)

Canary in the coal mine

Adjacent canary idea

Terminator canary
- Value hard to reproduce because it would tell the copy to stop
- StackGuard: 0x00 0D 0A FF
  - 0: String functions
  - newline: fgets(), etc.
  - -1 getc()
  - carriage return: similar to newline?
- Doesn’t stop: memcpy, custom loops

Random canary
- Can’t reproduce because attacker can’t guess
- For efficiency, usually one per execution
- Ineffective if disclosed
**XOR canary**
- Want to protect against non-sequential overwrites
- XOR return address with value $c$ at entry
- XOR again with $c$ before return
- Standard choice for $c$: see random canary

**Further refinements**
- More flexible to do earlier in compiler
- Rearrange buffers after other variables
  - Reduce chance of non-control overwrite
- Skip canaries for functions with only small variables
  - Who has an overflow bug in an 8-byte array?

**What’s usually not protected?**
- Backwards overflows
- Function pointers
- Adjacent structure fields
- Adjacent static data objects

**Where to keep canary value**
- Fast to access
- Buggy code/attacker can’t read or write
- Linux/x86: %gs:0x14

**Complex anti-canary attack**
- Canary not updated on fork in server
- Attacker controls number of bytes overwritten

**Complex anti-canary attack**
- Canary not updated on fork in server
- Attacker controls number of bytes overwritten
  - ANRY BNRY CNRY DNRY ENRY FNRY
  - search $2^{32} \rightarrow$ search $4 \cdot 2^8$
Shadow return stack
- Suppose you have a safe place to store the canary
- Why not just store the return address there?
- Needs to be a separate stack
- Ultimate return address protection

Outline
Return address protections
Announcements intermission
ASLR and counterattacks
W⊕X (DEP)

Note to early readers
- This is the section of the slides most likely to change in the final version
- If class has already happened, make sure you have the latest slides for announcements

Outline
Return address protections
Announcements intermission
ASLR and counterattacks
W⊕X (DEP)

Basic idea
- “Address Space Layout Randomization”
- Move memory areas around randomly so attackers can’t predict addresses
- Keep internal structure unchanged
  - E.g., whole stack moves together

Code and data locations
- Execution of code depends on memory location
- E.g., on 32-bit x86:
  - Direct jumps are relative
  - Function pointers are absolute
  - Data must be absolute
Relocation (Windows)
- Extension of technique already used in compilation
- Keep table of absolute addresses, instructions on how to update
- Disadvantage: code modifications take time on load, prevent sharing

PIC/PIE (GNU/Linux)
- “Position-Independent Code / Executable”
- Keep code unchanged, use register to point to data area
- Disadvantage: code complexity, register pressure hurt performance

What’s not covered
- Main executable (Linux 32-bit PIC)
- Incompatible DLLs (Windows)
- Relative locations within a module/area

Entropy limitations
- Intuitively, entropy measures amount of randomness, in bits
  Random 32-bit int: 32 bits of entropy
  ASLR page aligned, so at most $32 - 12 = 20$ bits of entropy
  Other constraints further reduce possibilities

Leakage limitations
- If an attacker learns the randomized base address, can reconstruct other locations
- Any stack address → stack unprotected, etc.

GOT hijack (Müller)
- Main program fixed, libc randomized
- PLT in main program used to call libc
- Rewire PLT to call attacker’s favorite libc functions
- E.g., turn printf into system
GOT hijack (Müller)

printf@plt: jmp *0x8049678
...

system@plt: jmp *0x804967c
...

0x8049678: <addr of printf in libc>
0x804967c: <addr of system in libc>

ret2pop (Müller)

- Take advantage of shellcode pointer already present on stack
- Rewrite intervening stack to treat the shellcode pointer like a return address
  - A long sequence of chained returns, one pop

ret2pop (Müller)

Outline

- Return address protections
- Announcements intermission
- ASLR and counterattacks
- W ⊕ X (DEP)

Basic idea

- Traditional shellcode must go in a memory area that is
  - writable, so the shellcode can be inserted
  - executable, so the shellcode can be executed
- But benign code usually does not need this combination
- W ⊕ X, really ¬(W ∧ X)

Non-writable code, X → ¬W

- E.g., read-only .text section
- Has been standard for a while, especially on Unix
- Lets OS efficiently share code with multiple program instances
Non-executable data, $W \rightarrow \neg X$

- Prohibit execution of static data, stack, heap
- Not a problem for most programs
  - Incompatible with some GCC features no one uses
  - Non-executable stack opt-in on Linux, but now near-universal

Implementing $W \oplus X$

- Page protection implemented by CPU
  - Some architectures (e.g., SPARC) long supported $W \oplus X$
  - x86 historically did not
    - One bit controls both read and execute
    - Partial stop-gap "code segment limit"
  - Eventual obvious solution: add new bit
    - NX (AMD), XD (Intel), XN (ARM)

One important exception

- Remaining important use of self-modifying code: just-in-time (JIT) compilers
  - E.g., all modern JavaScript engines
- Allow code to re-enable execution per-block
  - mprotect, VirtualProtect
  - Now a favorite target of attackers

Counterattack: code reuse

- Attacker can't execute new code
- So, take advantage of instructions already in binary
- There are usually a lot of them
- And no need to obey original structure

Classic return-to-libc (1997)

- Overwrite stack with copies of:
  - Pointer to libc’s `system` function
  - Pointer to `"/bin/sh"` string (also in libc)
- The `system` function is especially convenient
- Distinctive feature: return to entry point

Chained return-to-libc

- Shellcode often wants a sequence of actions, e.g.
  - Restore privileges
  - Allow execution of memory area
  - Overwrite system file, etc.
- Can put multiple fake frames on the stack
  - Basic idea present in 1997, further refinements
Beyond return-to-libc

- Can we do more? Oh, yes.
- Classic academic approach: what’s the most we could ask for?
- Here: “Turing completeness”
- How to do it: reading for Thursday

Next time

- Return-oriented programming (ROP)
  - And counter-defenses
- Control-flow integrity (CFI)