Vulnerability Assessment and Secure Programming

Yulei Sui

University of Technology Sydney, Australia

Software Vulnerability

Wiki's definition: A vulnerability is a weakness which can be exploited by a threat actor, such as an attacker, to cross privilege boundaries (i.e. perform unauthorized actions) within a computer system

Software Vulnerability

Wiki's definition: A vulnerability is a weakness which can be exploited by a threat actor, such as an attacker, to cross privilege boundaries (i.e. perform unauthorized actions) within a computer system

Causes of Software Vulnerabilities:

- High-level design flaws
- Insecure coding (This subject)

Common Types of Software Vulnerabilities

- Memory safety errors
 - Memory Leaks
 - Null pointer dereferences
 - Dangling pointers and use-after-frees
 - Buffer overflows
- Arithmetic errors
 - Integer overflows
 - Division by zero
- Tainted inputs
 - Tainted information flow
 - Code injection
 - Format string
 - SQL injection
- Side-channel attacks
 - Timing attacks

Let us take a look at examples of the above vulnerabilities, how to fix them and implement more secure programming practices (e.g., using assertions)

Memory Leaks (A Vulnerable Example)

A memory leak occurs when dynamically allocated memory is never freed along a program execution path.

Memory Leaks (A Vulnerable Example)

A memory leak occurs when dynamically allocated memory is never freed along a program execution path.

```
1 typedef struct Node {
                                          1 void free list(List *1) {
      int data:
                                                Node* current = 1->head:
      struct Node *next;
                                                while (current != NULL) {
4 } Node:
                                                     node *next = current->next;
5 typedef struct List {
                                                    delete current:
      struct Node *head:
                                                    current = next;
7 } List:
8
9 List* create_list(int num){
      List* list = new List():
10
                                         10 int main(){
      Node* current = list->head:
11
                                              List* 1 = create list(10): // create a list of 10 nodes
      for(i = 0; i < num; i++){</pre>
12
                                         12 // ... do something with the list
           current = new Node():
13
                                         free_list(1); // deallocate the memory for the list
14
           current = current->next:
                                         14 }
15
16
      return list:
17 }
```

Memory Leaks (A Vulnerable Example)

A memory leak occurs when dynamically allocated memory is never freed along a program execution path.

```
1 typedef struct Node {
                                          1 void free list(List *1) {
      int data:
                                                Node* current = 1->head:
      struct Node *next;
                                                while (current != NULL) {
4 } Node:
                                                    node *next = current->next;
5 typedef struct List {
                                                    delete current:
      struct Node *head:
                                                    current = next;
7 } List:
8
9 List* create_list(int num){
      List* list = new List():
10
                                         10 int main(){
      Node* current = list->head:
11
                                              List* 1 = create list(10): // create a list of 10 nodes
      for(i = 0; i < num; i++){</pre>
12
                                         12 // ... do something with the list
           current = new Node():
13
                                         13
                                              free_list(1); // deallocate the memory for the list
14
          current = current->next:
                                         14 }
15
16
      return list:
                                            Do we have a bug in the above code?
17 }
```

How do we fix the bua?

Memory Leaks (A Secure Example)

```
1 typedef struct Node {
                                          1 void free list(List *1) {
2
      int data:
      struct Node *next:
                                                Node* current = 1->head:
4 } Node:
                                                while (current != NULL) {
5 typedef struct List {
                                                    node *next = current->next:
       struct Node *head:
                                                     delete current:
7 } List;
                                                    current = next;
8
9 List* create list(int num){
                                                delete 1:
      List* list = new List():
10
11
      Node* current = list->head;
                                         10
      for(i = 0; i < num; i++){</pre>
                                         11 int main(){
12
13
           current = new Node():
                                              List* 1 = create list(10): // create a list of 10 nodes
           current = current->next:
                                         13 // ... do something with the list
14
15
                                         14 free_list(1); // deallocate the memory for the list
                                         15 }
16
      return list:
17 }
```

Memory Leaks (A Secure Example)

```
1 typedef struct Node {
2
       int data:
       struct Node *next:
4 } Node:
5 typedef struct List {
       struct Node *head:
7 } List;
8
9 List* create_list(int num){
       List* list = new List():
10
11
       Node* current = list->head:
       for(i = 0; i < num; i++){</pre>
12
13
           current = new Node():
14
           current = current->next:
15
16
       return list:
17 }
```

```
1 void free list(List *1) {
      Node* current = 1->head:
      while (current != NULL) {
          node *next = current->next;
          delete current:
          current = next:
      delete 1:
8
10
11 int main(){
    List* l = create_list(10): // create a list of 10 nodes
    // ... do something with the list
13
14 free_list(l); // deallocate the memory for the list
    l = nullptr: // avoid misuse later
15
16 }
```

Null Pointer Dereferences (A Vulnerable Example)

A NULL pointer dereference occurs when the application dereferences a pointer that it expects to be valid, but is NULL, typically causing a crash or exit.

```
1 struct Student{
2 int id:
                                                    1 int main(int argc, char **argv){
3 char* name:
                                                           int stuID:
4 } Student:
                                                           scanf( "%d%c", &stuID);
5 Student students[10] = {...};
                                                           Student* student = findStuRecord(stuID):
6
                                                           printf("%s\n", student->name);
  Student* findStuRecord(int sID){
                                                    6 }
       for(int i = 0; i < 10; i++){
8
           if(students[i].id == sID)
10
               return &students[i]:
11
12
      return nullptr:
13 }
```

Null Pointer Dereferences (A Vulnerable Example)

A NULL pointer dereference occurs when the application dereferences a pointer that it expects to be valid, but is NULL, typically causing a crash or exit.

```
1 struct Student{
2 int id:
3 char* name:
4 } Student:
5 Student students[10] = {...};
6
  Student* findStuRecord(int sID){
       for(int i = 0; i < 10; i++){
8
           if(students[i].id == sID)
10
               return &students[i]:
11
12
      return nullptr:
13 }
```

```
1 int main(int argc, char **argv){
2    int stuID;
3    scanf( "%d%c", &stuID);
4    Student* student = findStuRecord(stuID);
5    printf("%s\n", student->name);
6 }
```

Do we have a bug in the above code? If so, where should we put the assertion to stop the bug?

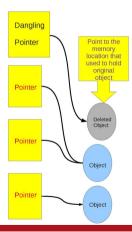
Null Pointer Dereferences (A Secure Example)

A NULL pointer dereference occurs when the application dereferences a pointer that it expects to be valid, but is NULL, typically causing a crash or exit.

```
1 struct Student{
2 int id:
3 char* name:
4 } Student:
                                                    1 int main(int argc, char **argv){
5 students[10] = {...};
                                                          int stuID:
                                                          scanf( "%d%c", &stuID):
6
  Student* findStuRecord(int sID){
                                                          Student* student = findStuRecord(stuID):
       for(int i = 0; i < 10; i++){
8
                                                          assert(student!=nullptr);
           if(students[i].id == sID)
                                                          printf("%s\n", student->name):
10
               return students[i]:
11
12
      return NULL:
13 }
```

Dangling references and use-after-frees

Dangling references are references that do not resolve to a valid memory object (e.g., caused by a use-after-free).



Dangling references and use-after-frees (A Vulnerable Example)

Dangling references are references that do not resolve to a valid memory object (e.g., caused by a use-after-free).

```
1 char* ptr = (char*)malloc(SIZE):
2 . . .
3 if (err) {
      abrt = 1;
      free(ptr);
8 if (abrt) {
       logError("operation aborted before commit", ptr);
10 F
```

Do we have a bug in the above code? If so, how should we fix the bug?

Dangling references and use-after-frees (A Vulnerable Example)

Dangling references and wild references are references that do not resolve to a valid destination

```
1 char* ptr = (char*)malloc(SIZE);
2 ...
3 if (err) {
4    abrt = 1;
5    free(ptr);
6    ptr = nullptr;
7 }
8    ...
9 if (abrt) {
10    assert(ptr != nullptr);
11    logError("operation aborted before commit", ptr);
12 }
```

Buffer Overflows

A buffer overflow occurs when the volume of data exceeds the storage capacity of the memory buffer, so that the program attempting to write the data to the buffer overwrites adjacent memory locations.



Buffer Overflows (A Vulnerable Example)

A buffer overflow occurs when the volume of data exceeds the storage capacity of the memory buffer, so that the program attempting to write the data to the buffer overwrites adjacent memory locations.

```
1 void bufferRead() {
      int n = 0;
     int ret = scanf("%d", &n):
      if (ret != 1 || n > 100) {
          return:
      char *p = (char *)malloc(n);
8
      int v = n:
      if (p == NULL) return:
      p[y] = 'a':
10
      free(p);
11
12 }
```

Do we have a bug in the above code? If so, where is the bug?

Buffer Overflows (A Secure Example)

A buffer overflow occurs when the volume of data exceeds the storage capacity of the memory buffer, so that the program attempting to write the data to the buffer overwrites adjacent memory locations.

```
1 void bufferRead() {
       int n = 0;
      int ret = scanf("%d", &n);
       if (ret != 1 | | n > 100) {
           return:
       char *p = (char *)malloc(n);
       int v = n:
       if (p == NULL) return:
       assert(y < n);</pre>
10
       p[y] = 'a';
11
       free(p):
12
13 }
```

An integer overflow occurs when an arithmetic operation attempts to create a numeric value that is outside of the range that can be represented with a given number of digits – either higher than the maximum or lower than the minimum representable value.

An integer overflow occurs when an arithmetic operation attempts to create a numeric value that is outside of the range that can be represented with a given number of digits either higher than the maximum or lower than the minimum representable value.

Typical binary register widths for unsigned integers include

- 4 bits: maximum representable value 2⁴ 1 = 15
- 8 bits: maximum representable value 28 1 = 255
- 16 bits: maximum representable value 2¹⁶ 1 = 65.535
- 32 bits: maximum representable value 2³² 1 = 4,294,967,295 (the most common width for personal computers as of 2005).
- 64 bits: maximum representable value 2^{64} 1 = 18,446,744,073,709,551,615 (the most common width for personal computer CPUs, as of 2017),
- 128 bits: maximum representable value 2¹²⁸ 1 = 340.282.366.920.938.463.463.374.607.431.768.211.455

In C/C++, overflows of **signed integer** will cause **undefined behavior** but overflows of unsigned integer will not. For unsigned integers, when the value exceeds the maximum value (2^n for some n), the result is reduced to that value modulo 2^n .

Here are some of C's various integer types and the values (in standard library):

- INT_MIN (the minimum value for an integer): -2,147,483,648 bits
- INT_MAX (the maximum value for an integer): +2,147,483,647 bits
- UINT_MAX (the maximum value for an unsigned integer): 4,294,967,295 bits

In C/C++, overflows of signed integer will cause undefined behavior but overflows of unsigned integer will not. For unsigned integers, when the value exceeds the maximum value (2^n for some n), the result is reduced to that value modulo 2^n .

Here are some of C's various integer types and the values (in standard library):

- INT_MIN (the minimum value for an integer): -2,147,483,648 bits
- INT_MAX (the maximum value for an integer): +2,147,483,647 bits
- UINT_MAX (the maximum value for an unsigned integer): 4,294,967,295 bits
- UINT MAX + 1 = ?
- UINT_MAX + 2 = ?
- UINT_MAX + 3 = ?

In C/C++, overflows of signed integer will cause undefined behavior but overflows of unsigned integer will not. The resulting unsigned integer type is reduced modulo to the number that is one greater than the largest value that can be represented by the resulting type (modulo power of two, i.e., 2^n where n is No. of bits).

Here are some of C's various integer types and the values (in standard library):

- INT_MIN (the minimum value for an integer): -2,147,483,648 bits
- INT_MAX (the maximum value for an integer): +2.147.483.647 bits
- UINT_MAX (the maximum value for an unsigned integer): 4,294,967,295 bits
- UINT MAX + 1 == 0
- IJINT MAX + 2 == 1
- UINT_MAX + 3 == 2

Integer Overflows¹

Compilers may exploit undefined behavior and optimize when there is an overflow of a signed integer.

¹ https://stackoverflow.com/questions/23889022/gcc-optimizations-based-on-integer-overflow

Integer Overflows¹

Compilers may exploit undefined behavior and optimize when there is an overflow of a signed integer.

```
1 signed int x ;
2 \text{ if}(x > x + 1)
                                                       1 if (INT MAX + 1 < 0){
3 {
                                                             // Overlooked
      // Handle potential overflow
                                                       3 }
5 }
```

Here since a signed integer overflow is not defined, compiler is free to assume that it may never happen and hence it can optimize away the "if" blocks.

```
1 signed i = 1:
2 while (i > 0){
      i *= 2:
```

The above behavior is undefined and when compiled under '-O3' using GCC (version 4.7.2 on Debian 4.7.2-5). It performs branch simplifications, producing an infinite loop.

https://stackoverflow.com/questions/23889022/gcc-optimizations-based-on-integer-overflow

Integer Overflows (A Vulnerable Example)

```
1 int nresp = packet_get_int();
2
3 if (nresp > 0) {
4  response = xmalloc(nresp*sizeof(char*));
5
6  for (i = 0; i < nresp; i++)
7   response[i] = packet_get_string(NULL);
8 }</pre>
```

Do we have a bug in the above code? If so, where is the bug?

Integer Overflows (A Vulnerable Example)

```
int nresp = packet_get_int();
if (nresp > 0) {
 response = xmalloc(nresp*sizeof(char*));
 for (i = 0: i < nresp: i++)
    response[i] = packet_get_string(NULL);
```

Do we have a bug in the above code? If so, where is the bug?

- An integer overflow that leads to a buffer overflow found in an older version of OpenSSH (3.3):
- If nresp is greater than or equal to 1073741824 and sizeof(char*) is 4 (which is typical of 32-bit systems), then nresp*sizeof(char*) results in an overflow. Therefore, xmalloc() receives and allocates a small buffer. The subsequent loop causes a heap buffer overflow, which may, in turn, be used by an attacker to execute arbitrary code.

Integer Overflows (A Secure Example)

```
int nresp = packet_get_int();
assert(nresp < userDefinedSize);
if (nresp > 0) {
  response = xmalloc(nresp*sizeof(char*));

for (i = 0; i < nresp; i++)
  response[i] = packet_get_string(NULL);
}</pre>
```

Make sure the size received is under a user budget.

Integer Overflows (A Secure Example)

```
int nresp = packet_get_int();
   assert(nresp < userDefinedSize);</pre>
  if (nresp > 0) {
   response = xmalloc(nresp*sizeof(char*));
5
   for (i = 0; i < nresp; i++)</pre>
    response[i] = packet_get_string(NULL);
8 }
```

Make sure the size received is under a user budget.

```
1 int nresp = packet_get_int();
3 if (nresp > 0) {
4 assert(nresp <= userDefinedSize/sizeof(char</pre>
  *));
  response = xmalloc(nresp*sizeof(char*));
  for (i = 0: i < nresp: i++)
    response[i] = packet_get_string(NULL);
```

Make sure the malloc can safely allocate memory under a user budget.

Division by Zero (A Vulnerable Example)

The product divides a value by zero. The C standard (C11 6.5.5) states that dividing by zero has undefined behavior for either integer or floating-point operands. It can cause program crash or miscompilation by compilers.

```
1 unsigned computeAverageResponseTime (unsigned totalTime, unsigned numRequests)
2 {
3     return totalTime / numRequests;
4 }
```

Is the above code secure?

Division by Zero (A Secure Example)

The product divides a value by zero. The C standard (C11 6.5.5) states that dividing by zero has undefined behavior for either integer or floating-point operands. It can cause program crash or miscompilation by compilers.

```
1 unsigned computeAverageResponseTime (unsigned totalTime, unsigned numRequests)
2 {
3     assert(numRequests > 0);
4    return totalTime / numRequests;
5 }
```

Tainted Information Flow (A Vulnerable Example)

Malicious user inputs may cause unexpected program behaviors, information leakage or attacks when executing a target program.

Tainted Information Flow (A Vulnerable Example)

Malicious user inputs may cause unexpected program behaviors, information leakage or attacks when executing a target program.

```
1 void main(int argc, char **argv)
2 {
      char *pMsg = packet get string():
      ParseMsg((LOGIN_MSG_BODY *)pMsg);
5 }
6
  void ParseMsg(LOGIN_MSG_BODY *stLoginMsgBody)
8 {
      for (int ulIndex = 0; ulIndex < stLoginMsgBody->usLoginPwdLen; ulIndex++) {
9
          // do something
10
11
12 }
```

Do we have a bug in the above code? If so, where is the bug?

Tainted Information Flow (A Secure Example)

```
1 void main(int argc, char **argv)
2 {
3
      char *pMsg = packet_get_string();
      ParseMsg((LOGIN_MSG_BODY *)pMsg);
5 }
  void ParseMsg(LOGIN_MSG_BODY *stLoginMsgBody)
8 {
       assert(stLoginMsgBody->usLoginPwdLen < userDefinedSize);</pre>
9
       for (int ulIndex = 0; ulIndex < stLoginMsgBody->usLoginPwdLen; ulIndex++) {
10
11
           // do something
12
13 }
```

User input can cause long loops to hang the target program.

Code Injection (A Vulnerable Example)

Code injection is the exploitation of a computer bug that is caused by processing invalid data. The result of successful code injection can be disastrous, for example, by disclosing confidential information.

```
string user_id, command;
command = "cat user_info/";
cin >> user_id; // user_id is an integer
system command + user_id);
```

Is the above code secure?

Code Injection (A Vulnerable Example)

Code injection is the exploitation of a computer bug that is caused by processing invalid data. The result of successful code injection can be disastrous, for example, by disclosing confidential information.

```
string user_id, command;
command = "cat user_info/";
cin >> user_id; // user_id is an integer
system command + user_id);
```

Is the above code secure?

- Before system(), we did not check the legality of user_id. first, if x is a piece of dangerous code, it will be executed in system() and return harmful results.
- For example, if user types "05 && ipconfig", it will disclose ip address.

Code Injection (A Secure Example)

To detect the bug, we need to make sure after appending \$user_id will not generate another command.

```
string user_id, command;
command = "cat user_info/";
cin>>user_id; // what if user input "05 && ipconfig"
assert(isdigit(user_id));
system(command + user_id);
```

In the assert call, we should check whether user_id consists entirely of numeric characters, so that another malicious command would not be injected.

The Format String exploit occurs when the submitted data of an input string is evaluated as a command by the application.

```
1 #include <stdio.h>
2 void main(int argc, char **argv){
3     printf("%s\n", argv[1]);
4     printf(argv[1]);
5 }
```

The Format String exploit occurs when the submitted data of an input string is evaluated as a command by the application.

```
1 #include <stdio.h>
2 void main(int argc, char **argv){
3     printf("%s\n", argv[1]);
4     printf(argv[1]);
5 }
```

- The line printf("%s", argv[1]); is safe, if you compile the program and run it:
 - ./example "Hello World %s%s%s%s%s"

The Format String exploit occurs when the submitted data of an input string is evaluated as a command by the application.

```
1 #include <stdio.h>
2 void main(int argc, char **argv){
3     printf("%s\n", argv[1]);
4     printf(argv[1]);
5 }
```

- The line printf("%s", argv[1]); is safe, if you compile the program and run it:
 - ./example "Hello World %s%s%s%s%s%s"
 - Output: "Hello World %s%s%s%s%s%s"

The Format String exploit occurs when the submitted data of an input string is evaluated as a command by the application.

```
1 #include <stdio.h>
2 void main(int argc, char **argv){
     printf("%s\n", argv[1]);
     printf(argv[1]);
```

- The line printf(argv[1]); in the example is vulnerable
- ./example "Hello World %s%s%s%s%s%s"
- Output: program crash.
- The printf in the second line will interpret the %s%s%s%s%s%s in the input string as a reference to string pointers. It will try to interpret every %s as a pointer to a string. which will get to an invalid address, and attempting to access it will likely cause the program to crash

The Format String exploit occurs when the submitted data of an input string is evaluated as a command by the application.

```
1 #include <stdio.h>
2 void main(int argc, char **argv){
     printf("%s\n", argv[1]);
     printf(argv[1]);
```

- It can be more than a crash! An attacker can also use this to get information. For example, running: ./example "Hello World %p %p %p %p %p"
- First printf output: Hello World %p %p %p %p %p %p
- Second printf output: Hello World 000E133E 000E133E 0057F000 CCCCCCCC CCCCCCC CCCCCCC
- Causing possible memory address leakage and exploit²

²https://cs155.stanford.edu/papers/formatstring-1.2.pdf

SQL Injection (A Vulnerable Example)

The attacker can add additional SQL statements at the end of the query statements defined in advance (e.g., in the web application).

```
1 txtUserId = getRequestString("UserId");
2 txtSQL = "SELECT * FROM Users WHERE UserId = " + txtUserId;
```

Is the above code vulnerable? Why?

SQL Injection (A Vulnerable Example)

The attacker can add additional SQL statements at the end of the query statements defined in advance (e.g., in the web application).

```
1 txtUserId = getRequestString("UserId");
2 txtSQL = "SELECT * FROM Users WHERE UserId = " + txtUserId;
```

Is the above code vulnerable? Why?

- The original purpose of the code was to create an SQL statement to select a user, with a given user id.
- If there is nothing to prevent a user from entering "wrong" input, the user can enter some "smart" input like this: "105 OR 1=1".
- "OR 1==1" will make logical expression after WHERE always true. Then the sql database will return all the items from the Users table.

SQL Injection(A Secure Example)

To detect the bug, we need to verify the user's input.

```
1 txtUserId = "105 OR 1=1";
2 assert(isdigit(txtUserId));
3 txtSQL = "SELECT * FROM Users WHERE UserId = " + txtUserId:
```

Assuming that user id is an integer, we should assert the user's input is exactly an integer.

Side-Channel Attacks (A Vulnerable Example)

```
1 bool insecureCmp(char *ca, char *cb, int length)
2 {
3    for (int i = 0; i < length; i++)
4        if (ca[i] != cb[i])
5        return false;
6    return true;
7 }</pre>
```

 The above C code demonstrates a typical insecure string comparison which stops testing as soon as a character doesn't match.

Side-Channel Attacks (A Vulnerable Example)

```
bool insecureCmp(char *ca, char *cb, int length)
2 {
   for (int i = 0; i < length; i++)
     if (ca[i] != cb[i])
        return false:
   return true:
```

- The above C code demonstrates a typical insecure string comparison which stops testing as soon as a character doesn't match.
- Assume a computer that takes 1 ms to check each character of a password using the above function to validate passwords.
- If the attacker guesses a completely wrong password, it will take 1 ms. Once they get one character right, it takes 2 ms. Two correct characters take 3 ms. and so on.

Side-Channel Attacks (A Secure Example)

```
1 bool insecureCmp(char *ca, char *cb, int length)
2 {
    for (int i = 0; i < length; i++)</pre>
      if (ca[i] != cb[i])
        return false;
    return true:
7 }
1 # String comparison time: 1ms
   'aaaaaaaaaaaaaa' == 'V1cHt2S67DADJIm9s'
3
  # String comparison time: 1ms
   'baaaaaaaaaaaaaaa' == 'V1cHt2S67DADJIm9s'
6
  # String comparison time: 2ms
   'Vaaaaaaaaaaaaaaa' == 'V1cHt2S67DADJIm9s'
10 # String comparison time: 3ms
11 'V1aaaaaaaaaaaaaaaa' == 'V1cHt2S67DADJIm9s'
12 # ...
```

Side-Channel Attacks (A Secure Example)

```
1 bool constTimeCmp(char *ca, char *cb, int length)
2 {
   bool result = true:
   for (int i = 0; i < length; i++)
     result &= (ca[i] == cb[i]):
   return result:
7 }
```

 The above secure version runs in constant-time by testing all characters and using a bitwise operation to accumulate the result.

Side-Channel Attacks (A Secure Example)

```
1 bool constTimeCmp(char *ca, char *cb, int length)
2 {
3   bool result = true;
4   int i;
5   for (i = 0; i < length; i++)
6     result &= (ca[i] == cb[i]);
7   assert(length == i); // make sure the comparison times is proportional to the length.
8   return result;
9 }</pre>
```

- Make sure the each time calling the comparison function always costs the same time (constant-time comparison)
- The above secure version runs in constant-time by testing all characters and using a bitwise operation to accumulate the result.

What's next?

- Start doing Assignment 1 including two sets of quizzes and a coding task and submit them before end of 14th March.
 - Read through the Assignment 1 instruction slides ('Assignment 1.pdf') on Canvas.
 - Coding task template and specification: https://github.com/SVF-tools/ Teaching-Software-Verification/wiki/Assignment-1