#### CS107, Lecture 16 Reverse Engineering Privacy and Trust + Heap Allocators

Reading: B&O 9.9, 9.11

 This document is copyright (C) Stanford Computer Science and Nick Troccoli, licensed under Creative Commons Attribution 2.5 License. All rights reserved.
 Based on slides created by Cynthia Lee, Chris Gregg, Jerry Cain, Lisa Yan and others.
 NOTICE RE UPLOADING TO WEBSITES: This content is protected and may not be shared, uploaded, or distributed. (without expressed written permission)

# gdb tips



layout split info reg

p \$eax

ni

si

p \$eflags

(ctrl-x a: exit, ctrl-l: resize, refresh: refresh, layout reg/asm, focus next)

View C, assembly, and gdb (lab5) Print all registers

Print register value Print all condition codes currently set

b \*0x400546

b \*0x400550 if \$eax > 98

Set breakpoint at assembly instruction Set conditional breakpoint

Next assembly instruction

Step into assembly instruction (will step into function calls)

# gdb tips



p/x \$rdi p/t \$rsi

x \$rdi x/4bx \$rdi x/4wx \$rdi

finish

Print register value in hex Print register value in binary

Examine the byte stored at this address Examine 4 bytes starting at this address Examine 4 ints starting at this address

Finish function, return to caller

#### assign5

You are a security researcher hired to explore potential vulnerabilities and issues at Stanford Bank. **3 core parts:** 

- 1. Uncovering ATM software vulnerabilities
- 2. Demonstrating how a data leak can lead to data aggregation and uncovering of personal information
- **3. Reverse engineering a secure program** discover 4 passwords needed to gain access to the system

#### Minivault

The **minivault** program is practice for part 3, SecureVault (it doesn't share code with SecureVault but is similar reverse-engineering practice).

You must provide correct passwords for 2 stages:

./minivault [stage1password] [stage2password]

**stage1** and **stage2** are 2 functions in minivault, each passed in the password for that stage. Our goal is to get both to return 1, and not 0.

## **Reverse Engineering Tips**

- **1.** Run the program live in GDB and step through. Reading and diagramming by hand is useful, but quickly becomes infeasible with larger programs.
- 2. Break the assembly into chunks
- 3. Use gdb to verify your hypotheses.
- **4. Document your knowns and unknowns.** Document and re-verify conflicting assumptions.
- 5. Use compiler explorer to see what code looks like in assembly.
- **6. Use library functions to your advantage.** If you spot a call to what looks like a library function, it's the real deal.
- 7. When tracing an unknown function, before dissecting its behavior first learn about the input/output of the function and what role it plays.

# **Demo: Minivault**

cp -r /afs/ir/class/cs107/lecture-code/lect16 .

## **Reverse Engineering Tips**

- **1.** Run the program live in GDB and step through. Reading and diagramming by hand is useful, but quickly becomes infeasible with larger programs.
- 2. Break the assembly into chunks
- 3. Use gdb to verify your hypotheses.
- **4. Document your knowns and unknowns.** Document and re-verify conflicting assumptions.
- 5. Use compiler explorer to see what code looks like in assembly.
- **6. Use library functions to your advantage.** If you spot a call to what looks like a library function, it's the real deal.
- 7. When tracing an unknown function, before dissecting its behavior first learn about the input/output of the function and what role it plays.

### Learning Goals

- Learn about the connections between privacy, security and trust
- Learn the restrictions, goals and assumptions of a heap allocator
- Understand the conflicting goals of utilization and throughput

#### **Lecture Plan**

- Privacy and Trust
- The heap so far
- What is a heap allocator?
- Heap allocator requirements and goals
- Method 0: Bump Allocator

#### **Lecture Plan**

- Privacy and Trust
- The heap so far
- What is a heap allocator?
- Heap allocator requirements and goals
- Method 0: Bump Allocator

### **Privacy and Trust**

- Our learning about assembly and program execution helps us better understand computer security (the protection of data, devices, and networks from disruption, harm, theft, unauthorized access or modification).
- Computer security is important in part because it enables privacy.
- In understanding computer security, it's essential to understand the context in which it comes up (privacy and trust).

#### **Data Breaches**

#### Privacy/trust example: data breaches

- California list of data security breaches: link
- How does a data breach make a customer feel?

What is privacy? 4 possible framings in two categories: **Individualist:** the value of privacy as an individual right

- Privacy as control of information controlling how our private information is shared with others.
- Privacy as autonomy capacity to choose/decide for ourselves what is valuable.

#### **Social:** the value of privacy for a group

- Privacy as **social good** social life would be unlivable without privacy.
- Privacy (protection) as based in **trust** privacy enables trusting relationships

Privacy as **control of information** – controlling how our information is communicated to others.

- Consent requires *free* choice with available alternatives and *informed* understanding of what is being offered.
- How many of you just skip past the terms of service for new online services you sign up for?
- Do you feel in control of your information with the services you choose to use? Why or why not? If you're working on a service, how can you respect privacy while achieving product goals?
- Control over personal data being collected (e.g. data exports from services you use, privacy dashboards, device privacy protections)

Privacy as **autonomy** – capacity to choose/decide for ourselves what is valuable.

- Links to autonomy over our own lives and our ability to lead them as we choose.
- Do you feel that your autonomy is always respected when using products and services? Why or why not?

"[P]rivacy is valuable because it acknowledges our respect for persons as autonomous beings with the capacity to love, care and like—in other words, persons with the potential to freely develop close relationships" (Innes 1992)

#### **Individualist Models of Privacy**

Privacy as **autonomy** and privacy as **control over information** focus the value of privacy at an individual level.

- Individual privacy can conflict with interests of society or the state.
- Many debates over "privacy vs. security" whether one should be sacrificed for the other
  - Apple v. FBI case re: unlocking iPhones (link)
  - Debates around encryption (<u>link</u>)
- Where do your beliefs fall in balancing privacy and security? When (if at all) is it ok to sacrifice one, and how much?

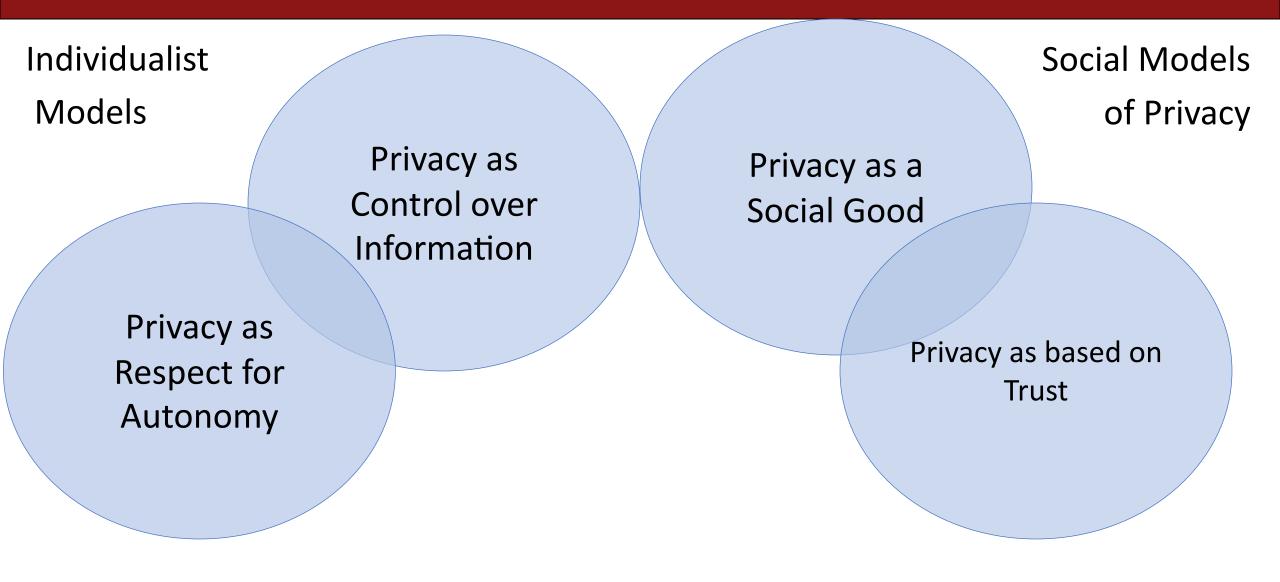
Privacy as **social good** – social life would be unlivable without privacy.

- Privacy has a social value in bringing about the kind of society we want to live in.
- What would society look like without privacy?

# Privacy (protection) as based in **trust** – privacy enables trusting relationships

- Privacy may help enable trusting relationships essential for cooperation.
  - For instance, a *fiduciary*: someone who stands in a legal or ethical relationship of trust with another person (or group). The fiduciary must act for the benefit of and in the best interest of the other person. E.g. tax filer with access to your bank account
    - Should anyone who has access to personal info have a *fiduciary* responsibility? (Richards & Hartzog 2020).
- This model of privacy stresses the essential relationship of trust placed in any holder of personal data and the responsibilities that result from this trust.

#### **Models of Privacy**



#### Who Should We Trust?

Both security and privacy rely on trusted people (who administer security, perform penetration tests, submit vulnerabilities to databases, or keep private information secret). The final piece of the security puzzle is understanding trust.

#### Trust = Reliance + Risk of Betrayal

What makes trust unique to relationships between people is that trust exposes one to being *betrayed or being let down* (Baier 1986).

#### **Penetration Testing & Trust**

**Penetration testing** is the practice of encouraging or hiring security researchers / contractors to find vulnerabilities in one's own code or system.

- Position of trust tester is given access to the system and encouraged to find exploitable vulnerabilities, expected to share what they have found with you.
- Means *relying on* their skill at finding vulnerabilities and *trusting* that their ethical compass will lead them to tell you and to act as a trustworthy *fiduciary* (guardian of your interests).

In Assignment5, you will have the opportunity to test your own ethical compass!

### Loss of Privacy

Loss of privacy can cause us various harms, including:

- **Aggregation:** combining personal information from various sources to build a profile of someone
- Exclusion: not knowing how our information is being used, or being unable to access or modify it (Google removing personal info from search – <u>link</u>)
- Secondary Use: using your information for purposes other than what was intended without permission.

#### **Mitigation: Differential Privacy**

- **Differential privacy** is a formal measure of privacy for datasets to try and protect individuals from aggregation by making them harder to identify (Dwork 2008).
- Imagine a large database, e.g., a medical database, with personal information and records of past activity tied to a name.
- The records might be useful for research purposes, or to train a machine learning model to predict future health outcomes, but what if giving access to the records exposed the privacy of individual person's health records?
- Differential privacy adds inconsequential noise (e.g., changing a birthday from 2001 to 2002) or removes records to make individuals harder to identify while preserving the utility of the dataset overall.

#### **Trust Models**

In every evaluation of privacy, we can ask: who is trusted? Who is distrusted? Does this model concentrate trust (and therefore power) in a single individual or small group, or does it distribute trust?

#### **Differential Privacy's Trust Model**

Differential privacy assumes that the only threat to privacy is an *external user querying the database* who must be prevented from aggregating data that could identify a user.

- In other words, the *trust model* of differential privacy is that the database owners and maintainers are to be fully trusted, and no one else.
- But is that the only threat? Differential privacy does not protect against improper use by people with full access to data or against leaks of the whole database, which may be the primary data exposure risks.

Differential privacy also does not question the assumption that amassing & storing large amounts of personal data is worth the risk of inevitable leaks (Rogaway 2015).

#### **Lecture Plan**

- Privacy and Trust
- The heap so far
- What is a heap allocator?
- Heap allocator requirements and goals
- Method 0: Bump Allocator

# How do the core malloc/realloc/free memory-allocation operations work?

### **CS107 Topic 6**

#### How do the core malloc/realloc/free memory-allocation operations work?

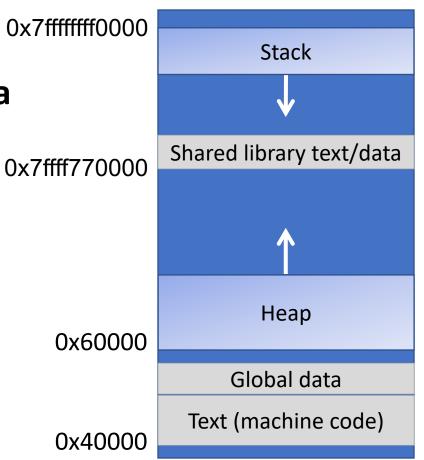
Why is answering this question important?

- Combines techniques from across the quarter (bits/bytes, pointers, memory, generics, assembly, efficiency, testing, and more) to understand a real-world system that you have relied on all quarter!
- Learning about the design and tradeoffs in a real-world large system gives us a great example of how to evaluate different designs when there's no one "right" answer.

**assign6:** implement two different possible designs for a heap allocator, implementing malloc/realloc/free.

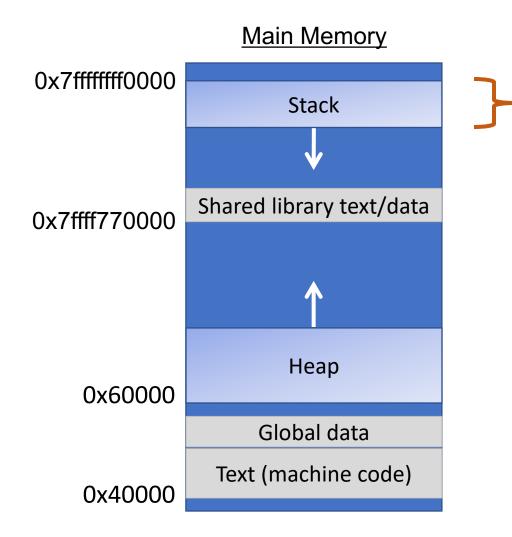
# Running a program

- Creates new process
- Sets up address space/segments
- Read executable file, load instructions, global data Mapped from file into gray segments
- Libraries loaded on demand
- Set up stack Reserve stack segment, init %rsp, call main
- malloc written in C, will init self on use Asks OS for large memory region, parcels out to service requests



#### **The Stack**



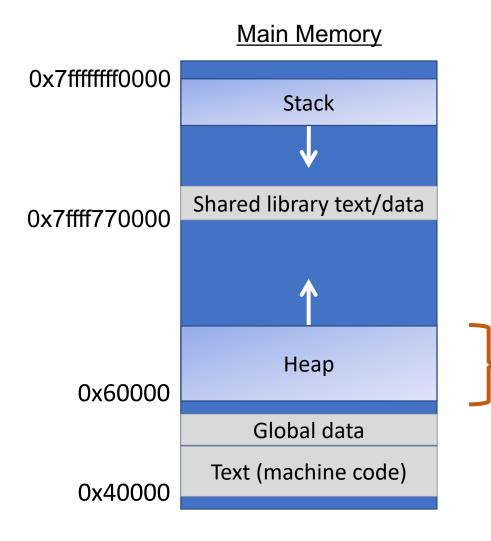


Stack memory "goes away" after function call ends.

Automatically managed at compile-time by gcc

From Assembly:
Stack management ==
moving %rsp around
(pushq, popq, mov)

#### **Today: The Heap**



Heap memory persists until caller indicates it no longer needs it.

Managed by C standard library functions (malloc, realloc, free)

<u>This lecture</u>: How does heap management work?

#### **Lecture Plan**

- Privacy and Trust
- The heap so far
- What is a heap allocator?
- Heap allocator requirements and goals
- Method 0: Bump Allocator

#### Your role so far: Client

void \*malloc(size\_t size);

Returns a pointer to a block of heap memory of at least size bytes, or NULL if an error occurred.

void free(void \*ptr);

Frees the heap-allocated block starting at the specified address.

#### void \*realloc(void \*ptr, size\_t size);

Changes the size of the heap-allocated block starting at the specified address to be the new specified size. Returns the address of the new, larger allocated memory region.

#### Your role now: Heap Hotel Concierge



(aka Heap Allocator)

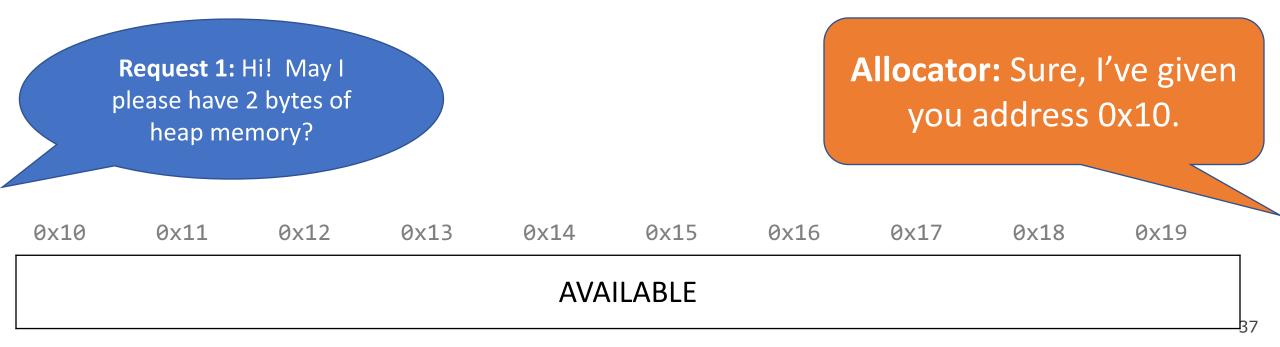
http://screencrave.com/wp-content/uploads/2014/03/the-grand-budapest-hotelanderson-image-2.jpg

#### What is a heap allocator?

- A heap allocator is a set of functions that fulfills requests for heap memory.
- On initialization, a heap allocator is provided the starting address and size of a large contiguous block of memory (the heap).

0x10	0x11	0x12	0x13	0x14	0x15	0x16	0x17	0x18	0x19
AVAILABLE									

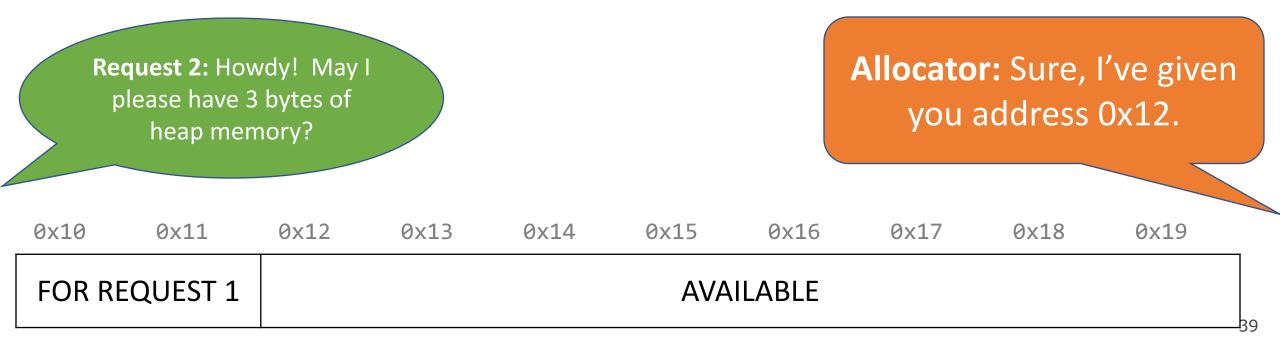
- A heap allocator is a set of functions that fulfills requests for heap memory.
- On initialization, a heap allocator is provided the starting address and size of a large contiguous block of memory (the heap).
- A heap allocator must manage this memory as clients request or no longer need pieces of it.



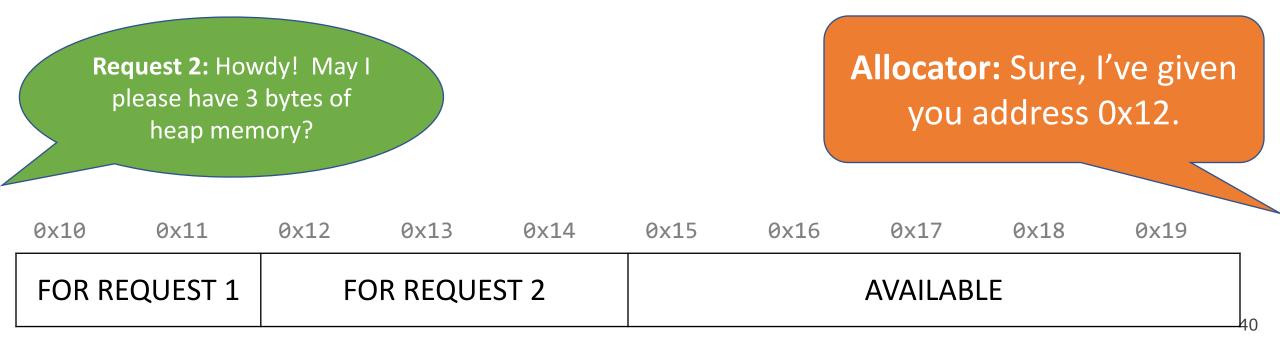
- A heap allocator is a set of functions that fulfills requests for heap memory.
- On initialization, a heap allocator is provided the starting address and size of a large contiguous block of memory (the heap).
- A heap allocator must manage this memory as clients request or no longer need pieces of it.



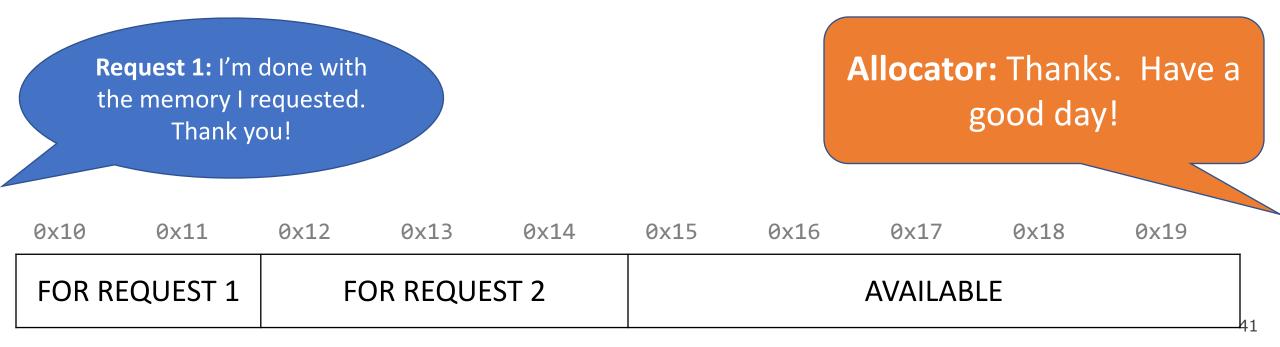
- A heap allocator is a set of functions that fulfills requests for heap memory.
- On initialization, a heap allocator is provided the starting address and size of a large contiguous block of memory (the heap).
- A heap allocator must manage this memory as clients request or no longer need pieces of it.



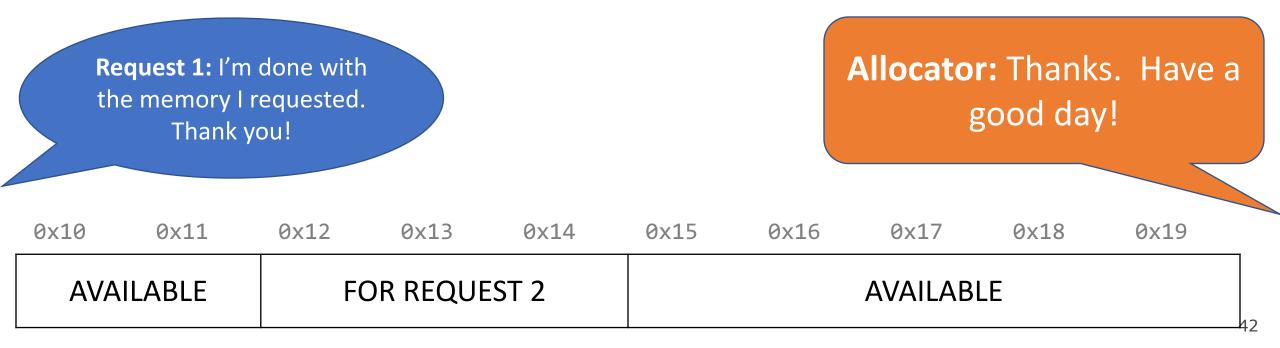
- A heap allocator is a set of functions that fulfills requests for heap memory.
- On initialization, a heap allocator is provided the starting address and size of a large contiguous block of memory (the heap).
- A heap allocator must manage this memory as clients request or no longer need pieces of it.



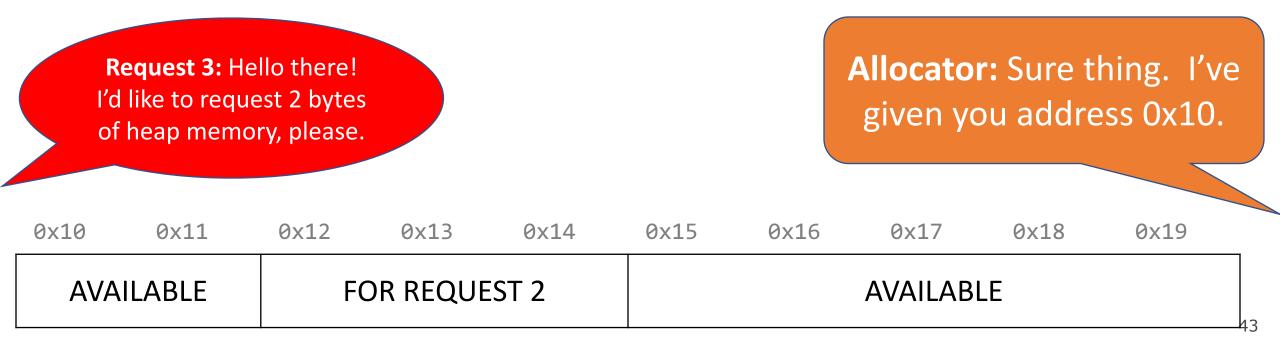
- A heap allocator is a set of functions that fulfills requests for heap memory.
- On initialization, a heap allocator is provided the starting address and size of a large contiguous block of memory (the heap).
- A heap allocator must manage this memory as clients request or no longer need pieces of it.



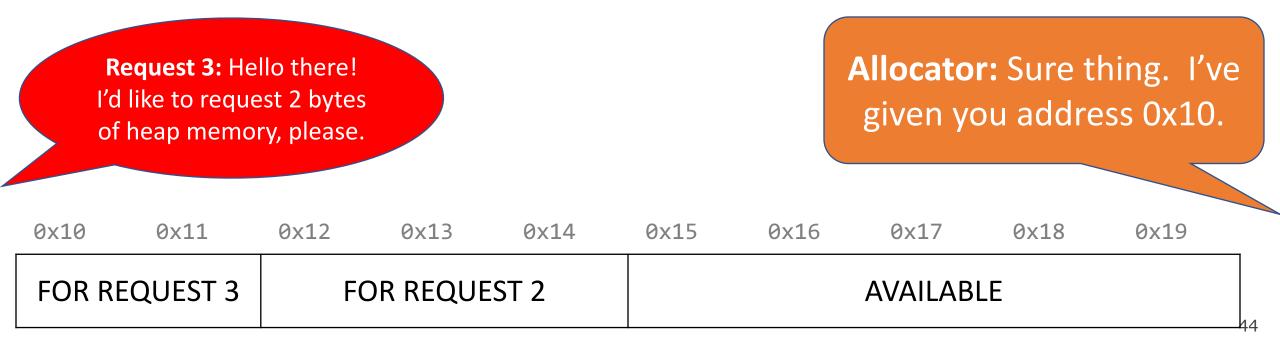
- A heap allocator is a set of functions that fulfills requests for heap memory.
- On initialization, a heap allocator is provided the starting address and size of a large contiguous block of memory (the heap).
- A heap allocator must manage this memory as clients request or no longer need pieces of it.



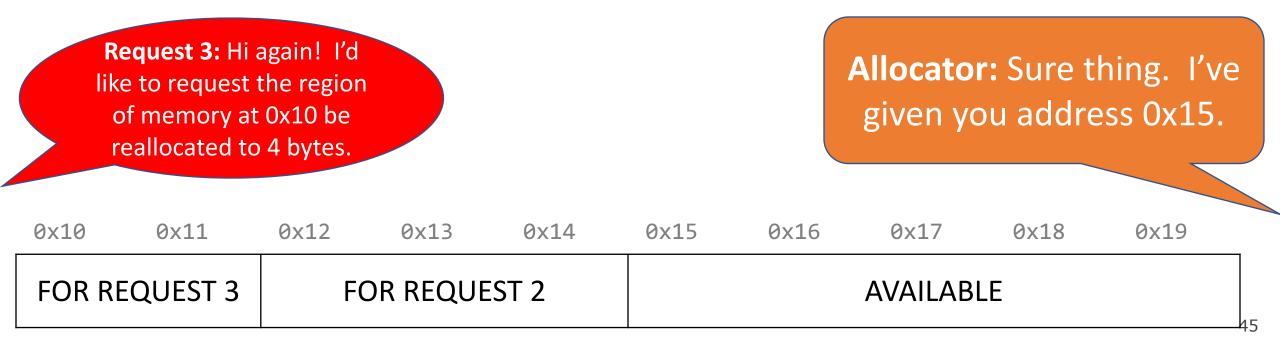
- A heap allocator is a set of functions that fulfills requests for heap memory.
- On initialization, a heap allocator is provided the starting address and size of a large contiguous block of memory (the heap).
- A heap allocator must manage this memory as clients request or no longer need pieces of it.



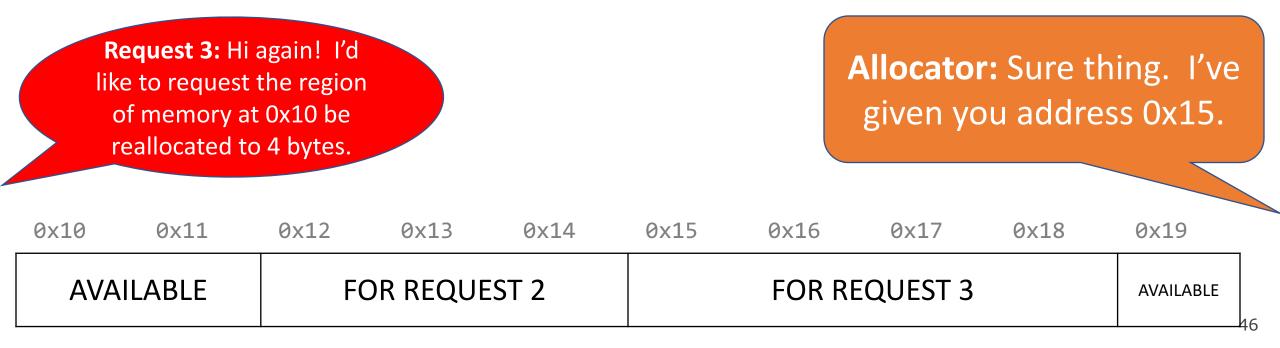
- A heap allocator is a set of functions that fulfills requests for heap memory.
- On initialization, a heap allocator is provided the starting address and size of a large contiguous block of memory (the heap).
- A heap allocator must manage this memory as clients request or no longer need pieces of it.



- A heap allocator is a set of functions that fulfills requests for heap memory.
- On initialization, a heap allocator is provided the starting address and size of a large contiguous block of memory (the heap).
- A heap allocator must manage this memory as clients request or no longer need pieces of it.



- A heap allocator is a set of functions that fulfills requests for heap memory.
- On initialization, a heap allocator is provided the starting address and size of a large contiguous block of memory (the heap).
- A heap allocator must manage this memory as clients request or no longer need pieces of it.



#### **Lecture Plan**

- Privacy and Trust
- The heap so far
- What is a heap allocator?
- Heap allocator requirements and goals
- Method 0: Bump Allocator

#### **Heap Allocator Functions**

void \*malloc(size\_t size);

void free(void \*ptr);

void \*realloc(void \*ptr, size\_t size);

A heap allocator must...

- 1. Handle arbitrary request sequences of allocations and frees
- 2. Keep track of which memory is allocated and which is available
- 3. Decide which memory to provide to fulfill an allocation request
- 4. Immediately respond to requests without delay
- 5. Return addresses that are 8-byte-aligned (must be multiples of 8).

A heap allocator must...

- 1. Handle arbitrary request sequences of allocations and frees
- 2. Keep track of which memory is allocated and which is available
- 3. Decide which memory to provide to fulfill an allocation request
- 4. Immediately respond to requests without delay
- 5. Return addresses that are 8-byte-aligned (must be multiples of 8).

A heap allocator cannot assume anything about the order of allocation and free requests, or even that every allocation request is accompanied by a matching free request.

A heap allocator must...

- 1. Handle arbitrary request sequences of allocations and frees
- 2. Keep track of which memory is allocated and which is available
- 3. Decide which memory to provide to fulfill an allocation request
- 4. Immediately respond to requests without delay
- 5. Return addresses that are 8-byte-aligned (must be multiples of 8).

A heap allocator marks memory regions as **allocated** or **available**. It must remember which is which to properly provide memory to clients.

A heap allocator must...

- 1. Handle arbitrary request sequences of allocations and frees
- 2. Keep track of which memory is allocated and which is available
- 3. Decide which memory to provide to fulfill an allocation request
- 4. Immediately respond to requests without delay
- 5. Return addresses that are 8-byte-aligned (must be multiples of 8).

A heap allocator may have options for which memory to use to fulfill an allocation request. It must decide this based on a variety of factors.

A heap allocator must...

- 1. Handle arbitrary request sequences of allocations and frees
- 2. Keep track of which memory is allocated and which is available
- 3. Decide which memory to provide to fulfill an allocation request
- 4. Immediately respond to requests without delay
- 5. Return addresses that are 8-byte-aligned (must be multiples of 8).

A heap allocator must respond immediately to allocation requests and should not e.g. prioritize or reorder certain requests to improve performance.

A heap allocator must...

- 1. Handle arbitrary request sequences of allocations and frees
- 2. Keep track of which memory is allocated and which is available
- 3. Decide which memory to provide to fulfill an allocation request
- 4. Immediately respond to requests without delay
- 5. Return addresses that are 8-byte-aligned (must be multiples of 8).

#### **Heap Allocator Goals**

- <u>Goal 1:</u> Maximize **throughput**, or the number of requests completed per unit time. This means minimizing the average time to satisfy a request.
- <u>Goal 2</u>: Maximize memory **utilization**, or how efficiently we make use of the limited heap memory to satisfy requests.

# Utilization

- The primary cause of poor utilization is **fragmentation**. **Fragmentation** occurs when otherwise unused memory is not available to satisfy allocation requests.
  - External Fragmentation (this example): no single space is large enough to satisfy a request, even though enough aggregate free memory is available
  - Internal Fragmentation: space allocated for a block is larger than needed (more later).
- In general: we want the largest address used to be as low as possible.

Request 6: Hi! May I

please have 4 bytes of

heap memory?

Allocator: I'm sorry, I don't have a 4 byte block available...

0x10	0x11	0x12	0x13	0x14	0x15	0x16	0x17	0x18	0x19	
Req. 1	Free	Req. 2	Free	Req. 3	Free	Req. 4	Free	Req. 5	Free	56

## Utilization

**Question**: Can we / should we shift these blocks down to make more space?

- YES, good idea!
- YES, but not a good idea for some reason
- NO, it can't be done!

0x10	0x11	0x12	0x13	0x14	0x15	0x16	0x17	0x18	0x19
Req. 1	Free	Req. 2	Free	Req. 3	Free	Req. 4	Free	Req. 5	Free
0x10	0x11	0x12	0x13	0x14	0x15	0x16	0x17	0x18	0x19
Req. 1	Req. 2	Req. 3	Req. 4	Req. 5			Free		
печ. т		neq. 5	NEY. 4	Neq. J			1166		

#### **Heap Allocator Goals**

- <u>Goal 1:</u> Maximize **throughput**, or the number of requests completed per unit time. This means minimizing the average time to satisfy a request.
- <u>Goal 2</u>: Maximize memory **utilization**, or how efficiently we make use of the limited heap memory to satisfy requests.

These are seemingly conflicting goals – for instance, it may take longer to better plan out heap memory use for each request. Heap allocators must find an appropriate balance between these two goals!

#### **Heap Allocator Goals**

- <u>Goal 1:</u> Maximize **throughput**, or the number of requests completed per unit time. This means minimizing the average time to satisfy a request.
- <u>Goal 2</u>: Maximize memory **utilization**, or how efficiently we make use of the limited heap memory to satisfy requests.

Other desirable goals: Locality ("similar" blocks allocated close in space) Robust (handle client errors) Ease of implementation/maintenance

#### **Lecture Plan**

- Privacy and Trust
- The heap so far
- What is a heap allocator?
- Heap allocator requirements and goals
- Method 0: Bump Allocator

Let's say we want to entirely prioritize throughput, and do not care about utilization at all. This means we do not care about reusing memory. How could we do this?

A **bump allocator** is a heap allocator design that simply allocates the next available memory address upon an allocate request and does nothing on a free request.

#### **Bump Allocator Performance**

# 1. Utilization

# 2. Throughput



 $\checkmark$ 

**Never** reuses memory

**Ultra fast**, short routines

- A **bump allocator** is a heap allocator design that simply allocates the next available memory address upon an allocate request and does nothing on a free request.
- Throughput: each **malloc** and **free** execute only a handful of instructions:
  - It is easy to find the next location to use
  - Free does nothing!
- Utilization: we use each memory block at most once. No freeing at all, so no memory is ever reused. ⊗
- We provide a bump allocator implementation as part of assign6 as a code reading exercise.

```
void *a = malloc(8);
void *b = malloc(4);
void *c = malloc(24);
free(b);
void *d = malloc(8);
```

0x10	0x14	0x18	0x1c	0x20	0x24	0x28	0x2c	0x30	0x34
				AVA	ILABLE				

void	*a =	<pre>malloc(8);</pre>	
void	*b =	<pre>malloc(4);</pre>	
void	*c =	<pre>malloc(24);</pre>	
free	(b);		
void	*d =	malloc(8);	

Variable	Value
а	0x10

0x10	0x14	0x18	0x1c	0x20	0x24	0x28	0x2c	0x30	0x34
	а				AVA	AILABLE			65

void	*a =	<pre>malloc(8);</pre>
void	*b =	<pre>malloc(4);</pre>
void	*c =	<pre>malloc(24);</pre>
free	(b);	
void	*d =	<pre>malloc(8);</pre>

Variable	Value
а	0x10
b	0x18

0x10	0x14	0x18	0x1c	0x20	0x24	0x28	0x2c	0x30	0x34
	а	b	+ padding	5		A	VAILABLE		66

void *a =	<pre>malloc(8);</pre>
void *b =	<pre>malloc(4);</pre>
<pre>void *c =</pre>	<pre>malloc(24);</pre>
<pre>free(b);</pre>	
void *d =	<pre>malloc(8);</pre>

Variable	Value
а	0x10
b	0x18
С	0x20

0x10	0x14	0x18	0x1c	0x20	0x24	0x28	0x2c	0x30	0x34
	а	b	+ padding	5			С		63

<pre>void *a = malloc(8);</pre>	Variable	Value
<pre>void *b = malloc(4);</pre>	а	0x10
<pre>void *c = malloc(24);</pre>		
<pre>free(b);</pre>	b	0x18
<pre>void *d = malloc(8);</pre>	С	0x20

0x10	0x14	0x18	0x1c	0x20	0x24	0x28	0x2c	0x30	0x34
a b + padding		5	С				68		

void	*a = m	alloc(	(8);				Variable		Value
	*b = m						а		0x10
void free(	*c = m b);	alloc(	24);				b		0x18
<pre>void *d = malloc(8);</pre>							C		0x20
							d		NULL
0x10	0x14	0x18	0x1c	0x20	0x24	0x28	0x2c	0x30	0x34

С

b + padding

а

## **Summary: Bump Allocator**

- A bump allocator is an extreme heap allocator it optimizes only for **throughput**, not **utilization**.
- Better allocators strike a more reasonable balance. How can we do this?

Questions to consider:

- 1. How do we keep track of free blocks?
- 2. How do we choose an appropriate free block in which to place a newly allocated block?
- 3. After we place a newly allocated block in some free block, what do we do with the remainder of the free block?
- 4. What do we do with a block that has just been freed?

#### Recap

- Privacy and Trust
- The heap so far
- What is a heap allocator?
- Heap allocator requirements and goals
- Method 0: Bump Allocator

Lecture 16 takeaways: Computer security comes up in discussions of privacy (individualist and social framings) and trust (reliance + risk of betrayal). How can we balance privacy and security? How can we mitigate potential harms?

A heap allocator is a set of functions that fulfills requests for heap memory. Seemingly-conflicting goals of maximizing throughput and memory utilization!