



# Chapter 10: Concurrent and Distributed Programming

## 10.2 Message-Passing and Graphical User Interface (GUI)

Xu Hanchuan  
[xhc@hit.edu.cn](mailto:xhc@hit.edu.cn)

May 29, 2019

# Outline

- **Sockets & Networking: Message passing between two computers**
  - Client/server design pattern
  - Network sockets and I/O
  - Using network sockets and wire protocols
  - Testing client/server code
- **Message passing with threads**
  - Implementing message passing with queues
  - Thread safety arguments with message passing
- **Graphical User Interface (GUI)**
  - View Tree
  - Model-View-Controller (MVC)
  - Background processing
- **Summary**



Recall: two models for concurrency



## Recall: two models for concurrency

- In the **shared memory model**, concurrent modules interact by reading and writing shared mutable objects in memory. 并发模块通过在内存中读写共享可变对象进行交互
  - Creating multiple threads inside a single Java process is our primary example of shared-memory concurrency.
- In the **message passing model**, concurrent modules interact by sending immutable messages to one another over a communication channel. 并发模块通过通信通道相互发送不可变消息进行交互
  - One example of message passing: the client/server pattern , in which clients and servers are concurrent processes, often on different machines, and the communication channel is a network socket .

# Advantages of message passing model

- **The message passing model has several advantages over the shared memory model, which boil down to greater safety from bugs.**
  - In message-passing, concurrent modules interact explicitly, by passing messages through the communication channel, rather than implicitly through mutation of shared data. 通过通信通道显式交互,而不是共享可变数据
  - The implicit interaction of shared memory can too easily lead to inadvertent interaction, sharing and manipulating data in parts of the program that don't know they're concurrent and aren't cooperating properly in the thread safety strategy.
  - Message passing also shares only immutable objects (the messages) between modules, whereas shared memory requires sharing mutable objects, which we have already seen can be a source of bugs . 消息传递共享的信息为不可变的, 降低了产生bugs的可能



# 1 Sockets & Networking: Message passing between processes of two computers



# Sockets & Networking

- *Client/server communication* over the network using the *socket* abstraction. 网络中的Client/server模式，采用socket作为抽象
- Network communication is inherently concurrent, so building clients and servers will require us to reason about their concurrent behavior and to implement them with thread safety. 网络通信是并发的，需要考虑线程安全问题
- We must also design the *wire protocol* that clients and servers use to communicate, just as we design the operations that clients of an ADT use to work with it. 需要设计client和server通讯用的线路协议
- Some of the operations with sockets are *blocking* : they block the progress of a thread until they can return a result. Blocking makes writing some code easier, but it also foreshadows a new class of concurrency bugs: deadlocks. socket的一些阻塞操作使代码编写更容易，但有可能造成死锁

# Client/server design pattern

- **Client/server design pattern** is for communication with message passing.
  - There are two kinds of processes: **clients and servers**. A client initiates the communication by connecting to a server. The client sends requests to the server, and the server sends replies back. Finally, the client disconnects. A server might handle connections from many clients concurrently, and clients might also connect to multiple servers. 客户端发起通信，服务器接收、处理、回复，重复此过程，客户端断开连接。服务器可同时处理多个客户端，客户端也可同时连接多个服务器。
- Many Internet applications work this way: web browsers are clients for web servers, an email program is a client for a mail server, etc.
- On the Internet, client and server processes are often running on different machines, connected only by the network, but it doesn't have to be that way — the server can be a process running on the same machine as the client. 客户机和服务器可在不同的计算机上，也可在一台计算机上。



# IP addresses

- **A network interface is identified by an IP address . IPv4 addresses are 32-bit numbers written in four 8-bit parts.**
  - 18.9.22.69 is the IP address of a MIT web server. Every address whose first octet is 18 is on the MIT network.
  - 18.9.25.15 is the address of a MIT incoming email handler.
  - 173.194.123.40 is the address of a Google web server.
  - **127.0.0.1 is the loopback or localhost address: it always refers to the local machine. Technically, any address whose first octet is 127 is a loopback address, but 127.0.0.1 is standard.**

# Hostnames

- **Hostnames are names that can be translated into IP addresses.**
- **A single hostname can map to different IP addresses at different times; and multiple hostnames can map to the same IP address.**
  - `www.hit.edu.cn` is the name for HIT's web server. You can translate this name to an IP address.
  - `dmz-mailsec-scanner-4.mit.edu` is the name for one of MIT's spam filter machines responsible for handling incoming email.
  - `google.com` is exactly what you think it is.
  - `localhost` is a name for `127.0.0.1` . When you want to talk to a server running on your own machine, talk to `localhost` .
- **Translation from hostnames to IP addresses is the job of the Domain Name System (DNS).**

# Port numbers

- A single machine might have multiple server applications that clients wish to connect to, so we need a way to direct traffic on the same network interface to different processes.
- Network interfaces have multiple ports identified by a 16-bit number from 0 to 65535.
- A server process binds to a particular port — it is now listening on that port. Clients have to know which port number the server is listening on. 服务器进程绑定到特定的端口，在该端口上进行侦听。客户端必须知道服务器正在侦听哪个端口号。
- When a client connects to a server, that outgoing connection also uses a port number on the client's network interface, usually chosen at random from the available non -well-known ports. 当客户端连接到服务器时，该连接还使用客户端网络接口上的端口号(通常从不常用端口中随机选择)

# Port numbers

- **There are some well-known ports which are reserved for system-level processes and provide standard ports for certain services.**
  - Port 22 is the standard SSH port. When you connect to athena.dialup.mit.edu using SSH, the software automatically uses port 22.
  - Port 25 is the standard email server port.
  - Port 80 is the standard web server port. When you connect to the URL `http://web.mit.edu` in your web browser, it connects to 18.9.22.69 on port 80.
- **When the port is not a standard port, it is specified as part of the address. For example, the URL `http://128.2.39.10:9000` refers to port 9000 on the machine at 128.2.39.10 .**

# Network sockets

- A **socket** is an endpoint in a network connection used to send and/or receive data **socket(套接字)**是用于发送和/或接收数据的网络连接中的端点（**socket本质是API，对TCP/IP协议的封装**）
- **Transport protocol:** TCP or UDP (or Raw IP, but not in Java)
- **Socket address:** local or remote IP address and port number
- **Sockets make network I/O feel like file I/O**
  - Support read, write, open, and close operations
  - Consistent with Unix philosophy “Everything’s a file.”
  - History: first appeared In Berkeley (BSD) Unix in 1983

# Network sockets

- **A socket represents one end of the connection between client and server.**

- A listening socket is used by a server process to wait for connections from remote clients. 服务器进程使用侦听套接字等待来自远程客户端的连接

In Java, use `ServerSocket` to make a listening socket, and use its `accept` method to listen to it.

- A connected socket can send and receive messages to and from the process on the other end of the connection. It is identified by both the local IP address and port number plus the remote address and port, which allows a server to differentiate between concurrent connections from different IPs, or from the same IP on different remote ports. 连接成功的套接字可以发送和接收来自连接另一端的进程的消息，通过本地和远程计算机的IP地址和端口号区分两端计算机

In Java, clients use a `Socket` constructor to establish a socket connection to a server. Servers obtain a connected socket as a `Socket` object returned from `ServerSocket.accept` .

# TCP networking in Java – java.net

## ■ IP Address – InetAddress

- `static InetAddress getByName(String host);`
- `static InetAddress getByAddress(byte[] b);`

## ■ Ordinary socket – Socket

- `Socket(InetAddress addr, int port);`
- `InputStream getInputStream();`
- `OutputStream getOutputStream();`
- `void close();`

## ■ Server socket – ServerSocket

- `ServerSocket(int port);`
- `Socket accept();`
- `void close();`
- ...

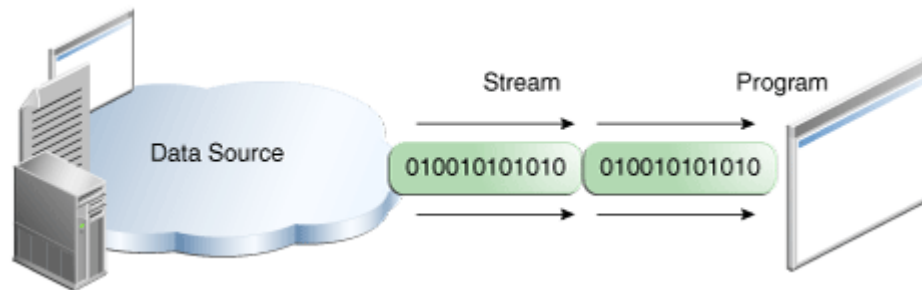
# I/O buffers

- The data that clients and servers exchange over the network is sent in chunks. 客户端和服务端通过网络交换的数据以块的形式发送
  - These are rarely just byte-sized chunks, although they might be.
  - The sending side (the client sending a request or the server sending a response) typically writes a large chunk (maybe a whole string like “HELLO, WORLD!” or maybe 20 megabytes of video data). 数据块可大可小
  - The network chops that chunk up into packets, and each packet is routed separately over the network. At the other end, the receiver reassembles the packets together into a stream of bytes. 网络将这些数据块分成数据包，每个数据包通过网络分别路由传输。另一端，接收器将数据包重新组装成一个字节流
- The result is a bursty kind of data transmission — the data may already be there when you want to read them, or you may have to wait for them to arrive and be reassembled. 有时需要等待数据到达重组
- When data arrive, they go into a **buffer**, an array in memory that holds the data until you read it. 当数据到达时，进入buffer，等待读取



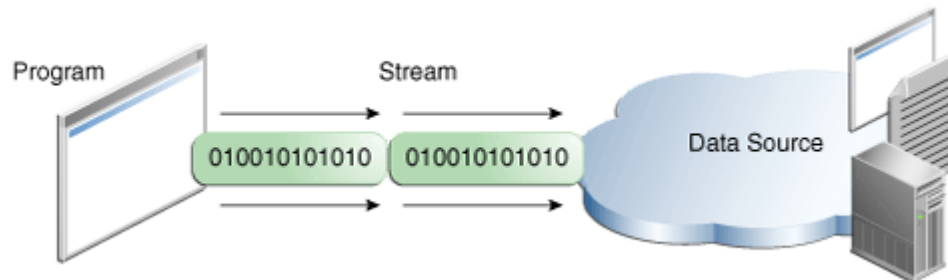
# I/O streams

- The data going into or coming out of a socket is a stream of bytes.
- In Java, `InputStream` objects represent sources of data flowing into your program:
  - Reading from a file on disk with a `FileInputStream`
  - User input from `System.in`
  - Input from a network socket



# I/O streams

- **OutputStream** objects represent data sinks, places we can write data to:
  - `FileOutputStream` for saving to files
  - `System.out` is a `PrintStream`, an `OutputStream` that prints readable representations of various types
  - Output to a network socket
- With sockets, remember that the *output* of one process is the *input* of another process. 使用sockets时，一个进程的输出是另一个进程的输入



# I/O blocking

- **Blocking means that a thread waits (without doing further work) until an event occurs.**
- **We can use this term to describe methods and method calls: if a method is a blocking method, then a call to that method can block , waiting until some event occurs before it returns to the caller.**
- **Socket input/output streams exhibit blocking behavior:套接字输入/输出流表现出阻塞行为:**
  - When an incoming socket's buffer is empty, calling read blocks until data are available. 当传入的套接字缓冲区为空时，读取操作被阻塞，直到数据可用
  - When the destination socket's buffer is full, calling write blocks until space is available. 当目标套接字缓冲区已满时，写入操作被阻塞，直到空间可用

# I/O blocking

- **Blocking is very convenient from a programmer's point of view, because the programmer can write code as if the read (or write ) call will always work, no matter what the timing of data arrival.**
  - If data (or for write , space) is already available in the buffer, the call might return very quickly.
  - If the read or write can't succeed, the call blocks. The operating system takes care of the details of delaying that thread until read or write can succeed.
- **Blocking happens throughout concurrent programming, not just in I/O (communication into and out of a process, perhaps over a network, or to/from a file, or with the user on the command line or a GUI, ...).**
- **Concurrent modules don't work in lockstep, like sequential programs do, so they typically have to wait for each other to catch up when coordinated action is required. 并发模块并不像顺序执行程序那样工作，所以当需要协调行动时，它们通常必须等待**

# Using network sockets

- Creating server- and client-side sockets and writing to and reading from their I/O streams.
- <http://docs.oracle.com/javase/tutorial/networking/sockets/index.html>

# Wire protocols

- Now that we have our client and server connected up with sockets, what do they pass back and forth over those sockets?
- A protocol is a set of messages that can be exchanged by two communicating parties. A protocol defines the rules syntax, semantics and synchronization of communication and possible error recovery methods. 协议规定了一组可由两个通信方进行交换的消息格式, 协议定义了通信的规则语法、语义和同步以及可能的错误恢复方法
- A wire protocol(线路协议) in particular is a set of messages represented as byte sequences, like hello world and bye (assuming we've agreed on a way to encode those characters into bytes).
- Most Internet applications use simple ASCII-based wire protocols.
  - HTTP
  - FTP
  - SMTP

# Designing a wire protocol

- When designing a wire protocol, apply the same rules of thumb you use for designing the operations of an abstract data type:
  - Keep the number of different messages small. It's better to have a few commands and responses that can be combined rather than many complex messages.
  - Each message should have a well-defined purpose and coherent behavior.
  - The set of messages must be adequate for clients to make the requests they need to make and for servers to deliver the results.
- Just as we demand representation independence from our types, we should aim for **platform-independence** in our protocols.
  - HTTP can be spoken by any web server and any web browser on any operating system.
  - The protocol doesn't say anything about how web pages are stored on disk, how they are prepared or generated by the server, what algorithms the client will use to render them, etc.

# Specifying a wire protocol

- **In order to precisely define for clients & servers what messages are allowed by a protocol, use a grammar.**

- For example, here is a very small part of the HTTP 1.1 request grammar from RFC 2616 section 5 :

```
request ::= request-line
          ((general-header | request-header | entity-header) CRLF)*
          CRLF
          message-body?
request-line ::= method SPACE request-uri SPACE http-version CRLF
method ::= "OPTIONS" | "GET" | "HEAD" | "POST" | ...
...
```

```
GET /aboutmit/ HTTP/1.1
Host: web.mit.edu
```

- GET is the method : we're asking the server to get a page for us.
- /aboutmit/ is the request-uri : the description of what we want to get.
- HTTP/1.1 is the http-version .
- Host: web.mit.edu is some kind of header.
- We don't have any message-body



# Testing client/server code

- **Concurrency is hard to test and debug. We can't reliably reproduce race conditions, and the network adds a source of latency that is entirely beyond our control. You need to design for concurrency and argue carefully for the correctness of your code.** 网络延迟增加了并发性测试和调试的难度，需要良好的设计确保正确性
- **(1) Separate network code from data structures and algorithms**
  - **Most of the ADTs in your client/server program don't need to rely on networking.** Make sure you specify, test, and implement them as separate components that are safe from bugs, easy to understand, and ready for change — in part because they don't involve any networking code. 客户/服务器程序中的大部分ADT都不需要依赖网络。确保将此类组件，单独规格说明、测试和实现，以利于理解、测试和修改。
  - If those ADTs will need to be used concurrently from multiple threads (for example, threads handling different client connections), our next reading will discuss your options. Otherwise, use the thread safety strategies of confinement, immutability, and existing threadsafe data types .

# Testing client/server code



## ■ (2) Separate socket code from stream code

- A function or module that needs to read from and write to a socket may only need access to the input/output streams, not to the socket itself.
- This design allows you to test the module by connecting it to streams that don't come from a socket.

# Summary of client/server design pattern

- **In the client/server design pattern, concurrency is inevitable: multiple clients and multiple servers are connected on the network, sending and receiving messages simultaneously, and expecting timely replies. 在客户端/服务器设计模式中，并发是不可避免的**
  - A server that blocks waiting for one slow client when there are other clients waiting to connect to it or to receive replies will not make those clients happy. **问题1：阻塞导致服务器对客户端的响应速度慢**
  - At the same time, a server that performs incorrect computations or returns bogus results because of concurrent modification to shared mutable data by different clients will not make anyone happy. **问题2：并发导致计算错误**
- **All the challenges of making our multi-threaded code safe from bugs , easy to understand , and ready for change apply when we design network clients and servers. 所有并发的挑战在C/S下都存在**
  - These processes run concurrently with one another (if on different machines), and any server that wants to talk to multiple clients concurrently (or a client that wants to talk to multiple servers) must manage that multi-threaded communication.



## 2 Message passing with threads



# Message passing with threads

- Message passing between processes: clients and servers communicating over network sockets. **Message passing**模型用于在客户端和服务端端的进程间通过sockets传递消息
- We can also use message passing between threads within the same process, and this design is often preferable to a shared memory design with locks. 在同一进程的线程间通过**message passing**传递消息，比通过锁定机制共享内存更受欢迎
- **Use a synchronized queue for message passing between threads.** The queue serves the same function as the buffered network communication channel in client/server message passing. 使用同步队列在线程之间传递消息

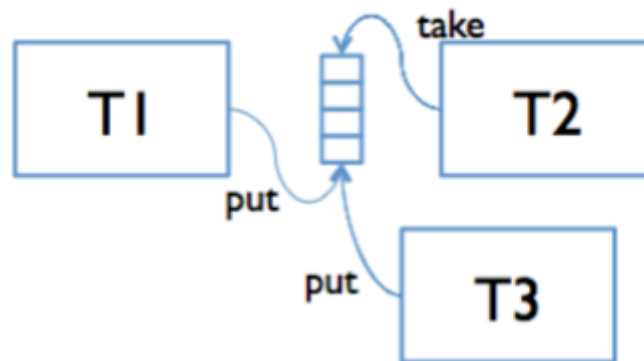
# How Java supports message passing

- Java provides the **BlockingQueue** interface for queues with blocking operations:
- In an ordinary Queue :
  - `add(e)` adds element `e` to the end of the queue.
  - `remove()` removes and returns the element at the head of the queue, throws an exception if the queue is empty.
- A **BlockingQueue** extends this interface:
  - additionally supports operations that wait for the queue to become non-empty when retrieving an element, and wait for space to become available in the queue when storing an element. 支持在检索元素时等待队列变为非空，在存储元素时等待队列中的空间变得可用。
  - `put(e)` blocks until it can add element `e` to the end of the queue (if the queue does not have a size bound, `put` will not block).
  - `take()` blocks until it can remove and return the element at the head of the queue, waiting until the queue is non-empty.

When you are using a **BlockingQueue** for message passing between threads, make sure to use the `put()` and `take()` operations, not `add()` and `remove()`.

# Producer-consumer design pattern

- Analogous to the client/server pattern for message passing over a network is the **producer-consumer design pattern** for message passing between threads.
  - Producer threads and consumer threads share a synchronized queue.
  - Producers put data or requests onto the queue, and consumers remove and process them.
  - One or more producers and one or more consumers might all be adding and removing items from the same queue. **This queue must be safe for concurrency.** 多个生产者和消费者共享一个同步的队列，都可对其写入和读取，需要同步安全机制



# Two implementations of BlockingQueue

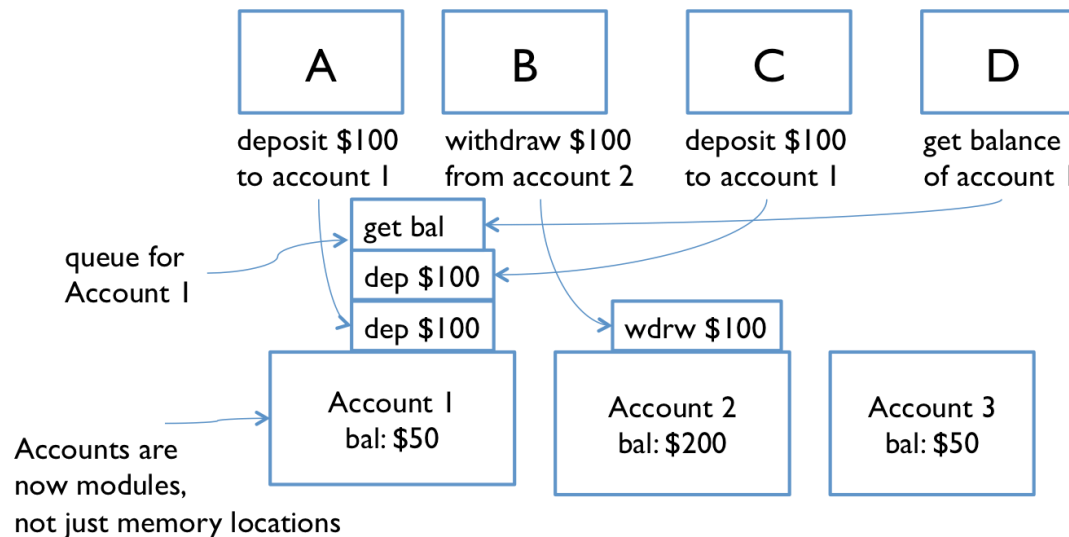
- Java provides two implementations of BlockingQueue :
  - **ArrayBlockingQueue** is a fixed-size queue that uses an array representation. putting a new item on the queue will block if the queue is full. 定长队列, 队列满时put操作会阻塞
  - **LinkedBlockingQueue** is a growable queue using a linked-list representation. If no maximum capacity is specified, the queue will never fill up, so put will never block. 可变长度队列
- Unlike the streams of bytes sent and received by sockets, these synchronized queues (like normal collections classes in Java) can hold objects of an arbitrary type. 可以保存任意类型的对象
  - Instead of designing a wire protocol, we must choose or design a type for messages in the queue. It must be an immutable type. 消息采用不可变类型
  - And just as we did with operations on a threadsafe ADT or messages in a wire protocol, we must design our messages here to prevent race conditions and enable clients to perform the atomic operations they need. 避免竞争情况和确保原子操作



# Bank account example

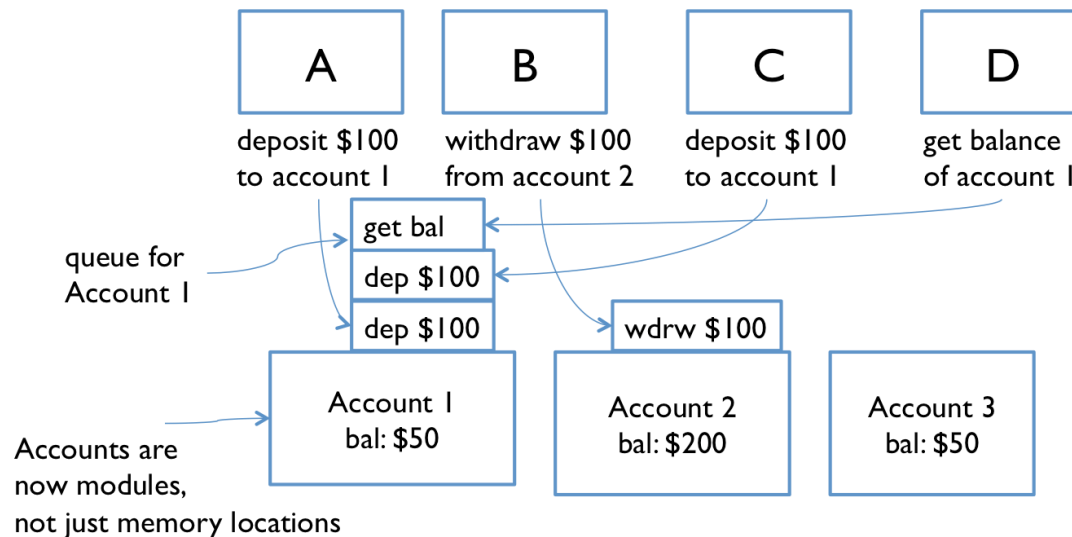
- Each cash machine and each account is its own module, and modules interact by sending messages to one another. Incoming messages arrive on a queue.
- We designed messages for get-balance and withdraw, and said that each cash machine checks the account balance before withdrawing to prevent overdrafts:

get-balance: if balance  $\geq$  1 then withdraw 1



# Bank account example

- But it is still possible to interleave messages from two cash machines so they are both fooled into thinking they can safely withdraw the last dollar from an account with only \$1 in it.
- We need to choose a better atomic operation: `withdraw-if-sufficient-funds` would be a better operation than just `withdraw`.



# Implementing message passing with queues

```

/** Squares integers. */
public class Squarer {

    private final BlockingQueue<Integer> in;
    private final BlockingQueue<SquareResult> out;
    // Rep invariant: in, out != null

    /** Make a new squarer.
     * @param requests queue to receive requests from
     * @param replies queue to send replies to */
    public Squarer(BlockingQueue<Integer> requests,
                   BlockingQueue<SquareResult> replies) {
        this.in = requests;
        this.out = replies;
    }

    /** Start handling squaring requests. */
    public void start() {
        new Thread(new Runnable() {
            public void run() {
                while (true) {
                    // TODO: we may want a way to stop the thread
                    try {
                        // block until a request arrives
                        int x = in.take();
                        // compute the answer and send it back
                        int y = x * x;
                        out.put(new SquareResult(x, y));
                    } catch (InterruptedException ie) {
                        ie.printStackTrace();
                    }
                }
            }
        }).start();
    }
}

```

Message passing  
module

```

/** An immutable squaring result message. */
public class SquareResult {
    private final int input;
    private final int output;

    /** Make a new result message.
     * @param input input number
     * @param output square of input */
    public SquareResult(int input, int output) {
        this.input = input;
        this.output = output;
    }

    @Override public String toString() {
        return input + "^2 = " + output;
    }
}

```

Outgoing message

```

public static void main(String[] args) {

    BlockingQueue<Integer> requests = new LinkedBlockingQueue<>();
    BlockingQueue<SquareResult> replies = new LinkedBlockingQueue<>();

    Squarer squarer = new Squarer(requests, replies);
    squarer.start();

    try {
        // make a request
        requests.put(42);
        // ... maybe do something concurrently ...
        // read the reply
        System.out.println(replies.take());
    } catch (InterruptedException ie) {
        ie.printStackTrace();
    }
}

```

Main() method that  
uses the squarer

# Stopping

- What if we want to shut down the Squarer so it is no longer waiting for new inputs? 如何关闭Squarer
  - In the client/server model, if we want the client or server to stop listening for our messages, we close the socket. 在C/S模式下, 可以关闭socket以停止客户端或者服务器继续侦听消息
  - If we want the client or server to stop altogether, we can quit that process. But here, the squarer is just another thread in the same process, and we can't "close" a queue. 如果期望服务器和客户端同时关闭, 可以退出进程
- One strategy is a poison pill: a special message on the queue that signals the consumer of that message to end its work.
  - To shut down the squarer, since its input messages are merely integers, we would have to choose a magic poison integer (everyone knows the square of 0 is 0 right? no one will need to ask for the square of 0...) or use null (don't use null). 一种不推荐的策略: 使用特殊的消息提示结束 (如0, 或者null), 但是魔数或者null都不是推荐的方式

# Stopping

- Instead, we might change the type of elements on the requests queue to an ADT:

$\text{SquareRequest} = \text{IntegerRequest} + \text{StopRequest}$

with **operations**:

$\text{input} : \text{SquareRequest} \rightarrow \text{int}$

$\text{shouldStop} : \text{SquareRequest} \rightarrow \text{boolean}$

and when we want to stop the squarer, we enqueue a StopRequest where shouldStop returns true.

方案1: 重新设计ADT, 使其具有结束标志和对应操作

# Stopping

```
public void run() {  
    while (true) {  
        try {  
            // block until a request arrives  
            SquareRequest req = in.take();  
            // see if we should stop  
            if (req.shouldStop()) { break; }  
            // compute the answer and send it back  
            int x = req.input();  
            int y = x * x;  
            out.put(new SquareResult(x, y));  
        } catch (InterruptedException ie) {  
            ie.printStackTrace();  
        }  
    }  
}
```

# interrupt()

- It is also possible to interrupt a thread by calling its `interrupt()` method. **方案2:采用`interrupt()`**
- If the thread is blocked waiting, the method it's blocked in will throw an `InterruptedException` (that's why we have to try-catch that exception almost any time we call a blocking method).
- If the thread was not blocked, an interrupted flag will be set.
- The thread must check for this flag to see whether it should stop working.
- **`interrupted()`检测当前线程是否被中断, 如果被中断, 返回`true`, 并清除中断标志**

```
public void run() {  
    // handle requests until we are interrupted  
    while ( ! Thread.interrupted()) {  
        try {  
            // block until a request arrives  
            int x = in.take();  
            // compute the answer and send it back  
            int y = x * x;  
            out.put(new SquareResult(x, y));  
        } catch (InterruptedException ie) {  
            // stop  
            break;  
        }  
    }  
}
```

# message-passing deadlock

- E.g., instead of using `LinkedBlockingQueues` that can grow arbitrarily (limited only by the size of memory), we will use the `ArrayBlockingQueue` implementation that has a fixed capacity:

```
private static final int QUEUE_SIZE = 100;
...
// make request and reply queues big enough to hold QUEUE_SIZE messages each
BlockingQueue<Integer> requests = new ArrayBlockingQueue<>(QUEUE_SIZE);
BlockingQueue<SquareResult> replies = new ArrayBlockingQueue<>(QUEUE_SIZ
E);
```

- Many message-passing systems use fixed-capacity queues for performance reasons, so this is a common situation. 采用定长队列可以提升性能



# message-passing deadlock

- To create the conditions needed for deadlock, the client code will make N requests, to get the squares of the numbers from 1 to N, before checking for any of Squarer's replies. Here is the full code:

```
private static final int QUEUE_SIZE = 100;
private static final int N = 100;

/** Use a Squarer to square all the integers from 1 to N. */
public static void main(String[] args) throws IOException {
    // make request and reply queues big enough to hold QUEUE_SIZE messages each
    BlockingQueue<Integer> requests = new ArrayBlockingQueue<>(QUEUE_SIZE);
    BlockingQueue<SquareResult> replies = new ArrayBlockingQueue<>(QUEUE_SIZE);

    Squarer squarer = new Squarer(requests, replies);
    squarer.start();

    try {
        // send the requests to square 1...N
        for (int x = 1; x <= N; ++x) {
            requests.put(x);
            System.out.println(x + "^2 = ?");
        }
        // collect the replies
        for (int x = 1; x <= N; ++x) {
            System.out.println(replies.take());
        }
    } catch (InterruptedException ie) {
        ie.printStackTrace();
    }
}
```

# message-passing deadlock

- As  $N$  grows larger and larger (far greater than 100), our client is making many requests without reading any replies.
- If  $N$  is larger than `QUEUE_SIZE`, the replies queue fills up with unread replies. Then Squarer blocks trying to put one more reply into that queue, and it stops calling `take` on the `requestsqueue`. The client can continue putting more requests into the requests queue, but only up to the size of that queue.
- If there are more additional requests than can fit in that queue – i.e., when  $N$  is greater than  $2 \times \text{QUEUE\_SIZE}$  – then the client's call to `requests.put()` will block too.
- And now we have our deadly embrace. Squarer is waiting for the client to read some replies and free up space on the replies queue, but the client is waiting for Squarer to accept some requests and free up space on the requests queue. Deadlock.

# Thread safety arguments with message passing

- **A thread safety argument with message passing might rely on:**
  - **Existing threadsafe data types for the synchronized queue.** This queue is definitely shared and definitely mutable, so we must ensure it is safe for concurrency. 队列本身是线程安全的数据类型
  - **Immutability** of messages or data that might be accessible to multiple threads at the same time. 传送的数据是不可变的
  - **Confinement** of data to individual producer/consumer threads. Local variables used by one producer or consumer are not visible to other threads, which only communicate with one another using messages in the queue. 生产者和消费者内部的数据对外部是不可见的，仅通过消息通讯
  - **Confinement** of mutable messages or data that are sent over the queue but will only be accessible to one thread at a time. This argument must be carefully articulated and implemented. But if one module drops all references to some mutable data like a hot potato as soon as it puts them onto a queue to be delivered to another thread, only one thread will have access to those data at a time, precluding concurrent access. 如果消息是可变的，则要确保同一时刻只有一个线程能够访问

# Final suggestions for preventing deadlock

- One solution to deadlock is to design the system so that there is no possibility of a cycle — so that if A is waiting for B, it cannot be that B was already (or will start) waiting for A. 一个方案是系统设计层面确保不会出现死锁
- Another approach to deadlock is timeouts. If a module has been blocked for too long (maybe 100 milliseconds? or 10 seconds? how to decide?), then you stop blocking and throw an exception. Now the problem becomes: what do you do when that exception is thrown? 另外一个方案是终止阻塞并抛出异常

# Thread safety arguments with message passing

- **Existing threadsafe data types** for the synchronized queue. This queue is definitely shared and definitely mutable, so we must ensure it is safe for concurrency.
- **Immutability** of messages or data that might be accessible to multiple threads at the same time.
- **Confinement** of data to individual producer/consumer threads. Local variables used by one producer or consumer are not visible to other threads, which only communicate with one another using messages in the queue.

# Thread safety arguments with message passing

- **Confinement** of mutable messages or data that are sent over the queue but will only be accessible to one thread at a time. This argument must be carefully articulated and implemented. Suppose one thread has some mutable data to send to another thread. If the first thread drops all references to the data like a hot potato as soon as it puts them onto a queue for delivery to the other thread, then only one thread will have access to those data at a time, precluding concurrent access. 限制可变消息或通过队列发送的数据，但一次只能由一个线程访问。实施时要格外注意。
- In comparison to synchronization, message passing can make it easier for each module in a concurrent system to maintain its own thread safety invariants. We don't have to reason about multiple threads accessing shared data if the data are instead transferred between modules using a threadsafe communication channel. 消息传递机制比同步机制更加容易维护线程安全的不变性，不需要推理多线程访问共享数据的问题，只要确保通讯通道的线程安全即可

# Summary

- Rather than synchronize with locks, message passing systems synchronize on a shared communication channel, e.g. a stream or a queue. 相比通过锁进行同步，消息传递同步是依靠消息通道进行共享
- Threads communicating with blocking queues is a useful pattern for message passing within a single process. 通过阻塞队列的方式是单进程内部线程间通信的有用方式



# 3 Graphical User Interfaces (GUI)





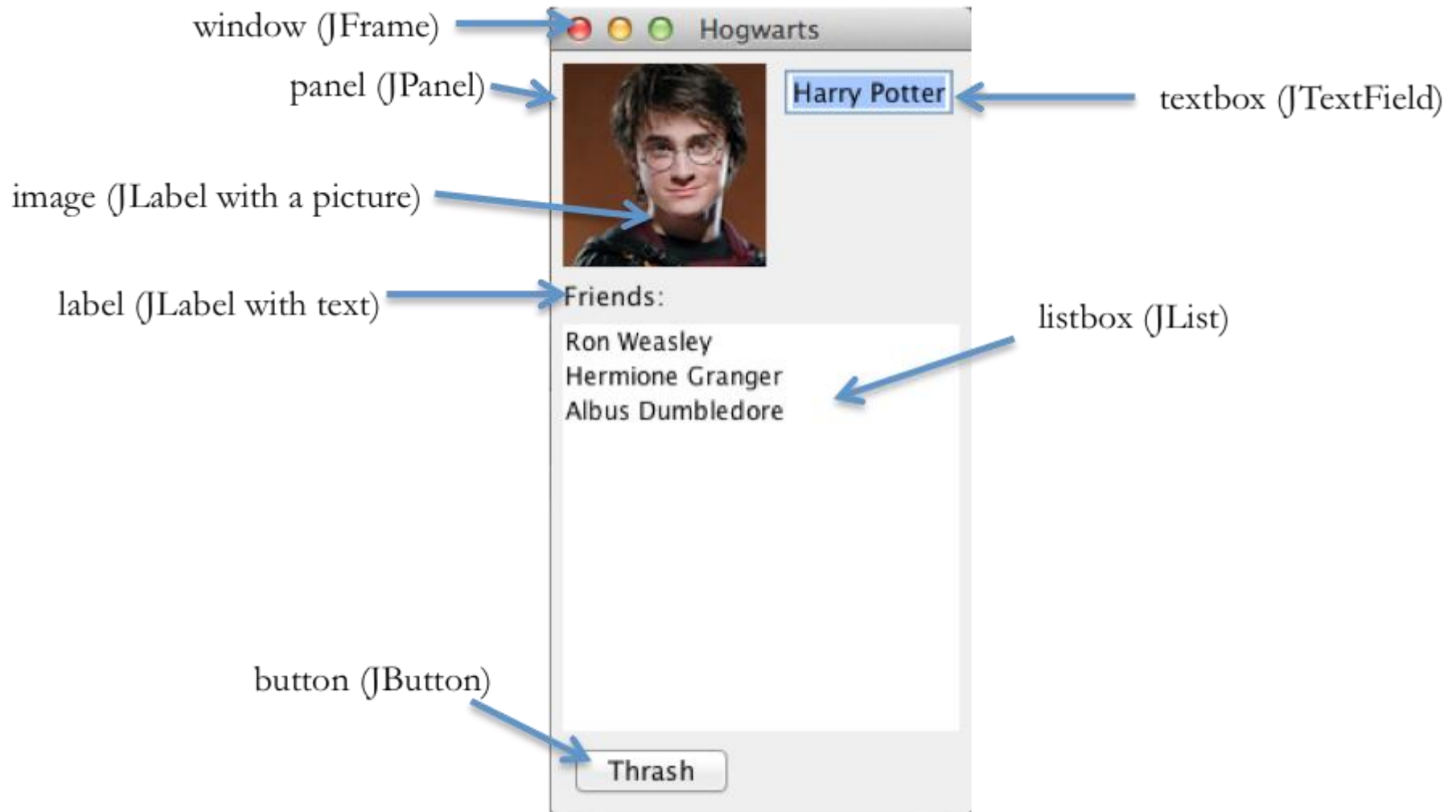
# Objectives

- To take a high-level look at the **software architecture of GUI software**, focusing on the design patterns that have proven most useful.
- **Three of the most important patterns:**
  - **View tree**, which is a central feature in the architecture of every important GUI toolkit;
  - **Model-view-controller pattern**, which separates input, output, and data;
  - **Listener pattern**, which is essential to decoupling the model from the view and controller.

# View Tree

- GUI are composed of view objects, each of which occupies a certain portion of the screen, generally a rectangular area called its bounding box.
- The view concept goes by a variety of names in various UI toolkits.
  - In Java Swing , they're JComponent objects;
  - In HTML, they're elements or nodes;
  - In other toolkits, they may be called widgets, controls, or interactors.
- View tree: views are arranged into a hierarchy of containment, in which some views contain other views.
  - Typical containers are windows, panels, and toolbars.
- The view tree is not just an arbitrary hierarchy, but is in fact a spatial one: child views are nested inside their parent's bounding box.

# View Tree



# How the view tree is used: (1) Output

- **Output. Views are responsible for displaying themselves, and the view tree directs the display process.**
- **GUIs change their output by mutating the view tree.**
  - For example, to show a new set of photos in a photo album GUI, the current thumbnails are removed from the view tree and a new set of thumbnails is added in their place.
  - A redraw algorithm built into the GUI toolkit automatically redraws the affected parts of the subtree.
    - In Java Swing, every view in the tree has a `paint()` method that knows how to draw itself on the screen.
    - The repaint process is driven by calling `paint()` on the root of the tree, which recursively calls `paint()` down through all the descendent nodes of the view tree.

## How the view tree is used: (2) Input

- **Input: views can have input handlers, and the view tree controls how mouse and keyboard input is processed.**
- Input is handled somewhat differently in GUIs than we've been handling it in parsers and servers.
  - In those systems, we've seen a single parser (**input loop**) that reads commands typed by the user or messages sent by the client, parses them, and decides how to direct them to different modules of the program.
  - If a GUI were written that way, it might look like this (in pseudocode):

```
while (true) {  
    read mouse click  
    if (clicked on Thrash button) doThrash();  
    else if (clicked on textbox) doPlaceCursor();  
    else if (clicked on a name in the listbox) doSelectItem();  
    ...  
}
```

# Listener pattern

- In a GUI, we don't directly write this kind of method, because it's not modular – it mixes up responsibilities for button, listbox, and textbox all in one place.
- Instead, GUIs exploit the spatial separation provided by the view tree to provide functional separation as well.
- Mouse clicks and keyboard events are distributed around the view tree, depending on where they occur.
- GUI input event handling is an instance of the **Listener pattern** (also known as Publish-Subscribe):
  - An event source generates a stream of discrete events, which correspond to state transitions in the source.
  - One or more listeners register interest (subscribe) to the stream of events, providing a function to be called when a new event occurs.

# Listener pattern

- In this case, the mouse is the event source, and the events are changes in the state of the mouse: its x,y position or the state of its buttons (whether they are pressed or released).
- Events often include additional information about the transition (such as the x,y position of mouse), which might be bundled into an event object or passed as parameters.
- When an event occurs, the event source distributes it to all subscribed listeners, by calling their callback methods.

```
JButton playButton = new JButton("Play");
```

```
playButton.addActionListener(new ActionListener() {  
    public void actionPerformed(ActionEvent event) {  
        playSound();  
    }  
});
```

# Listener pattern

- The control flow through a GUI proceeds like this:
  - A top-level event loop reads input from mouse and keyboard. In Java Swing, and most graphical user interface toolkits, this loop is actually hidden from you. It's buried inside the toolkit, and listeners appear to be called magically.
  - For each input event, it finds the right view in the tree (by looking at the x,y position of the mouse) and sends the event to that view's listeners.
  - Each listener does its thing (which might involve e.g. modifying objects in the view tree), and then *returns immediately to the event loop* .



# Listener pattern: ActionListener

- **Many GUI objects generate their own higher-level events, often as a result of some combination of low-level input events. For example:**
  - JButton sends an action event when it is pressed (whether by mouse or keyboard)
  - JList sends a selection event when the selected element changes (whether by mouse or by keyboard)
  - JTextField sends change events when the text inside it changes for any reason
- **A button can be pressed either by the mouse (with a mouse down and mouse up event) or by the keyboard (which is important for people who can't use a mouse, like blind users).**
  - Always listen for these high-level events, not the low-level input events.
  - Use an ActionListener to respond to a JButton press, not a mouse listener.

# How the view tree is used: (3) Layout


## ■ Layout

- The view tree controls how the views are laid out on the screen, i.e. how their bounding boxes are assigned.
- An automatic layout algorithm automatically calculates positions and sizes of views.
- Specialized containers (like `JSplitPane`, `JScrollPane`) do layout themselves.
- More generic containers ( `JPanel` , `JFrame` ) delegate layout decisions to a layout manager (e.g. `GroupLayout` , `BorderLayout` , `BoxLayout` , ...).

# Separating Frontend from Backend

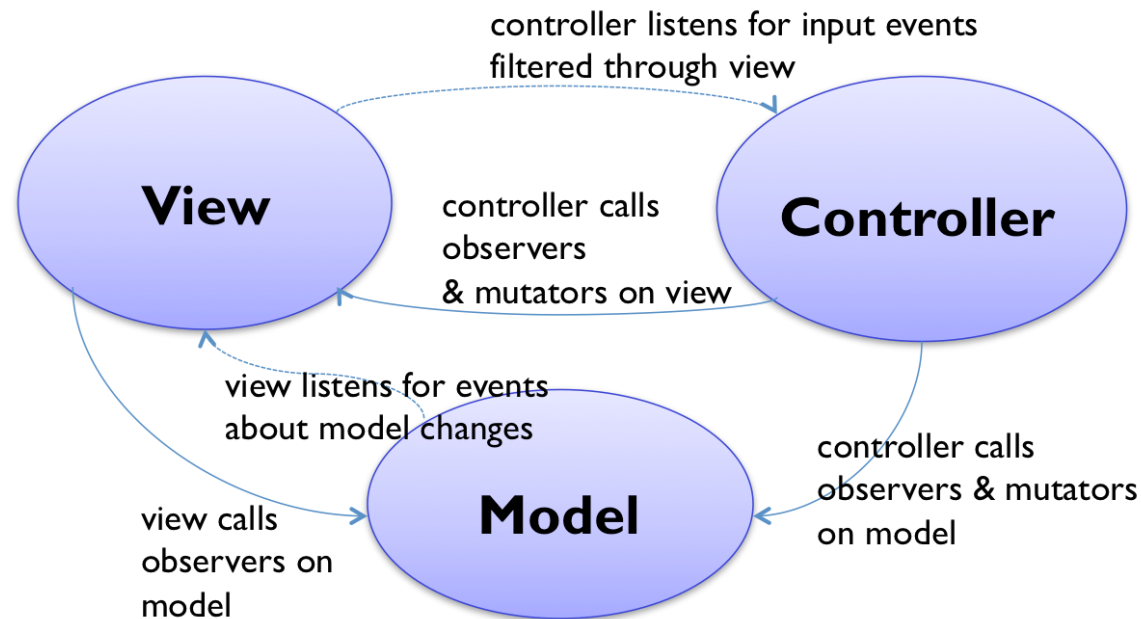
- We've seen how GUI programs are structured around a view tree, and how input events are handled by attaching listeners to views. This is the start of a separation of concerns – **output handled by views, and input handled by listeners.**
- But we're still missing the application itself – **the backend that represents the data and logic** that the user interface is showing and editing.
- Why do we want to separate this from the user interface?

# Callbacks

- 
- The actionPerformed listener is an example of a general design pattern, a callback.
  - A callback is a function that a client provides to a module for the module to call.
  - This is in contrast to normal control flow, in which the client is doing all the calling: calling down into functions that the module provides.
  - With a callback, the client is providing a piece of code for the implementer to call.

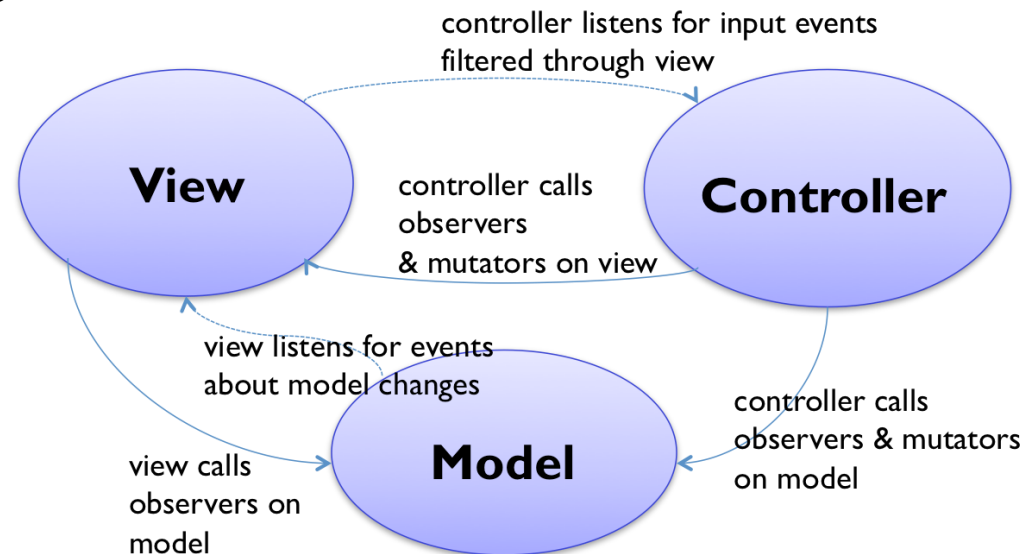
# Model-View-Controller (MVC) pattern

- The **Model-View-Controller pattern** has this separation of concerns as its primary goal.
- It **separates the user interface frontend from the application backend**, by putting backend code into the model and frontend code into the view and controller.
- **MVC also separates input from output**; the controller is supposed to handle input, and the view is supposed to handle output.



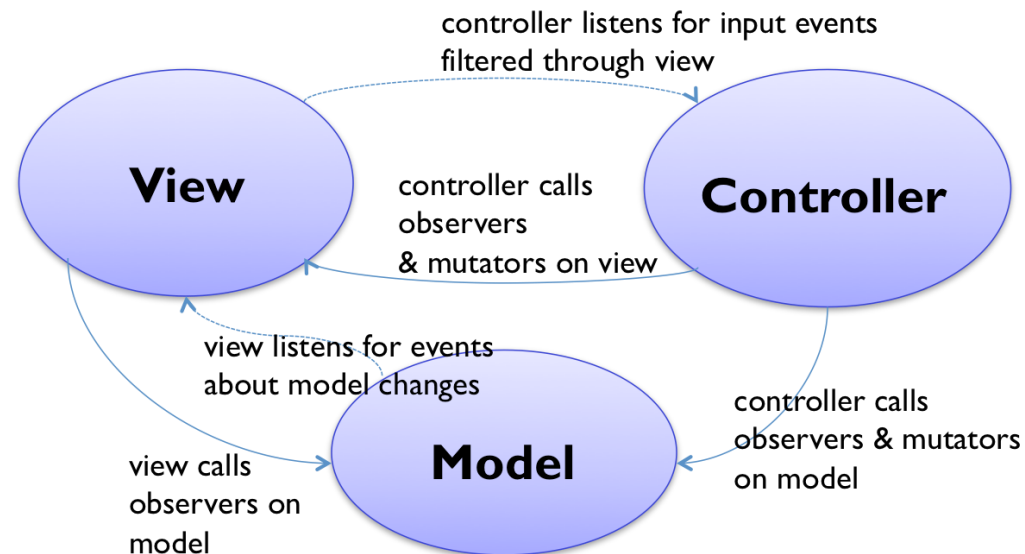
# Model

- The **model** is responsible for maintaining application-specific data and providing access to that data.
  - Models are often mutable, and they provide methods for changing the state safely, preserving its representation invariants.
  - But a model must also notify its clients when there are changes to its data, so that dependent views can update their displays, and dependent controllers can respond appropriately.
  - Models do this notification using the listener pattern, in which interested views and controllers register themselves as listeners for change events generated by the model.



# View and Controller

- **View** objects are responsible for output.
  - A view occupies some chunk of the screen, usually a rectangular area.
  - Basically, the view queries the model for data and draws the data on the screen. It listens for changes from the model so that it can update the screen to reflect those changes.
- **Controller** handles the input.
  - It receives keyboard and mouse events, and instructs the model to change accordingly.

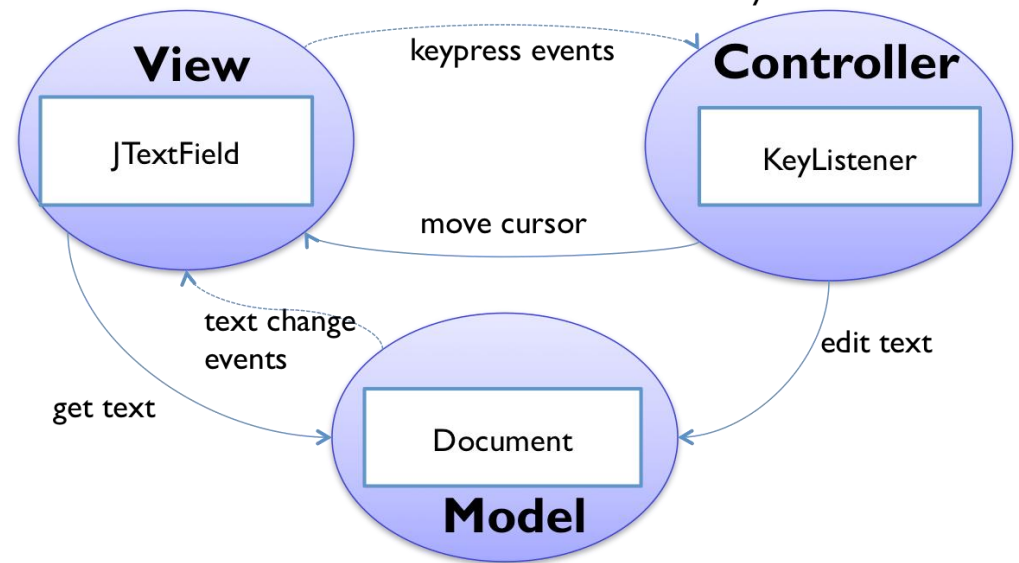


# A simple example of MVC

- A simple example of the MVC pattern is a text field.
- **TextField:**
  - Its model is a mutable string of characters.
  - The view is an object that draws the text on the screen (usually with a rectangle around it to indicate that it's an editable text field).
  - The controller is an object that receives keystrokes typed by the user and inserts them into the mutable string.

TextField is a JComponent that can be added to a view tree

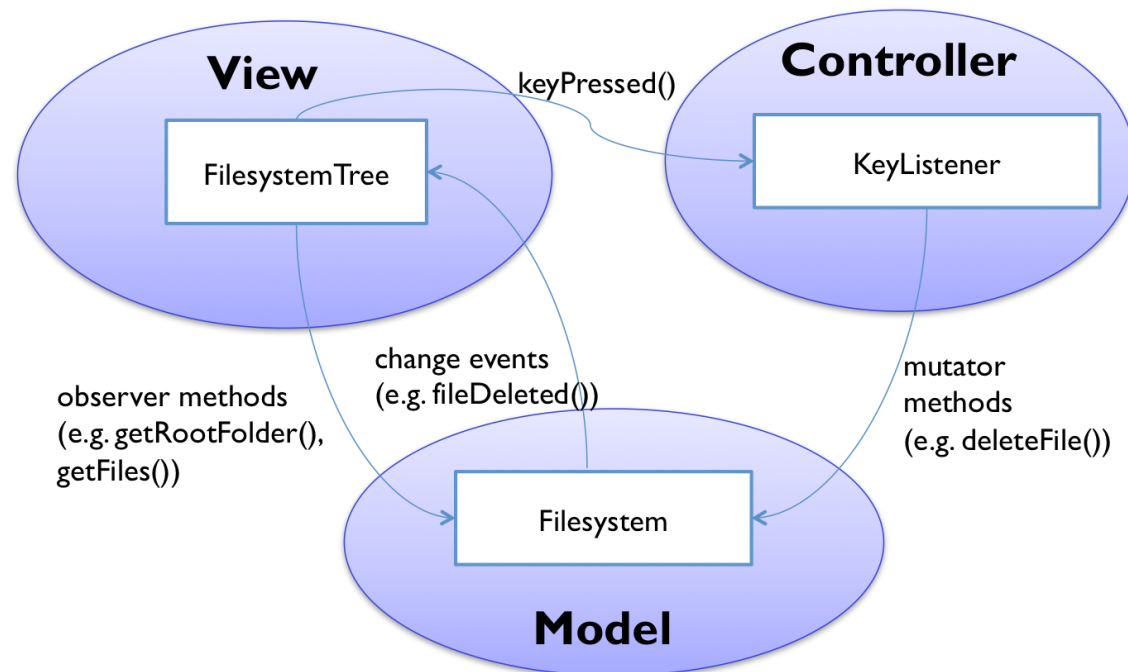
KeyListener is a listener for keyboard events





# A simple example of MVC

- Instances of the MVC pattern appear at many scales in GUI software.
  - At a higher level, this text field might be part of a view (like an address book editor), with a different controller listening to it (for text-changed events), for a different model (like the address book).
- Here's a larger example, in which the view is a filesystem browser, the model is the disk filesystem, and the controller is an input handler that translates the user's keystrokes and mouse clicks into operations on the model and view.



# Benefits of MVC

- **It allows the interface to have multiple views showing the same application data.**
  - For example, a database field might be shown in a table and in an editable form at the same time.
- **It allows views and models to be reused in other applications.**
  - The MVC pattern enables the creation of user interface toolkits, which are libraries of reusable views.
  - Java Swing is such a toolkit. You can easily reuse view classes from this library (like JButton and JTree ) while plugging your own models into them.

# Background Processing in GUI

- Why do we need to do background processing in GUI?
  - Even though computer systems are steadily getting faster, we're also asking them to do more.
  - Many programs need to do operations that may take some time: retrieving URLs over the network, running database queries, scanning a filesystem, doing complex calculations, etc.
- But GUIs are event-driven programs, which means everything is triggered by an input event handler.
  - For example, in a web browser, clicking a hyperlink starts loading a new web page.
  - But if the click handler is written so that it actually retrieves the web page itself, then the web browser will be very painful to use. ---- Because its interface will appear to freeze up until the click handler finishes retrieving the web page and returns to the event loop.

# What's the reason?

- This happens because input handling and screen repainting is all **handled from a single thread**.
  - That thread (called the event-dispatch thread) has a loop that reads an input event from the queue and dispatches it to listeners on the view tree.
  - When there are no input events left to process, it repaints the screen.
- But if an input handler you've written delays returning to this loop – because it's blocking on a network read, or because it's searching for the solution to a big Sudoku puzzle – then **input events stop being handled, and the screen stops updating**.
- So long tasks need to run in the background!

# Java GUI program is multithreaded

- **In Java, the event-dispatch thread is distinct from the main thread of the program.**
  - It is started automatically when a user interface object is created.
  - As a result, every Java GUI program is automatically multithreaded.
  - Many programmers don't notice, because the main thread typically doesn't do much in a GUI program – it starts creation of the view, and then the main thread just exits, leaving only the event-dispatch thread to do the main work of the program.
  
- **The fact that Swing programs are multithreaded creates risks.**
  - There's very often a shared mutable datatype in your GUI: the model.
  - If you use background threads to modify the model without blocking the event-dispatch thread, then you have to make sure your data structure is threadsafe.

# Java GUI program is multithreaded

- **Another important shared mutable datatype in your GUI is the view tree.**
  - Java Swing's view tree is not threadsafe: you cannot safely call methods on a Swing object from anywhere but the event-dispatch thread.
- **The view tree is a big meatball of shared state, and the Swing specification doesn't guarantee that there's any lock protecting it.**
  - Instead the view tree is confined to the event-dispatch thread , by specification.
  - So it's ok to access view objects from the event-dispatch thread (i.e., in response to input events), but the Swing specification forbids touching – reading or writing – any JComponent objects from a different thread.
  - In the actual Swing implementation, there is one big lock ( `Component.getTreeLock()` ) but only some Swing methods use it, so it's not effective as a synchronization mechanism.

# A safe way to access the view tree

- The safe way to access the view tree is to do it from the event-dispatch thread.
- Swing takes a clever approach: it uses the event queue itself as a message-passing queue, i.e., you can put your own custom messages on the event queue, the same queue used for mouse clicks, keypresses, button action events, and so forth.
- Your custom message is actually a piece of executable code, an object that implements `Runnable`, and you put it on the queue using `SwingUtilities.invokeLater`.

```
SwingUtilities.invokeLater(new Runnable() {  
    public void run() {  
        content.add(thumbnail);  
        ...  
    }  
});
```

# Summary of GUI

- The **view tree** organizes the screen into a tree of nested rectangles, and it is used in dispatching input events as well as displaying output.
- The **Listener pattern** sends a stream of events (like mouse or keyboard events, or button action events) to registered listeners.
- The **Model-View-Controller** pattern separates responsibilities: model=data, view=output, controller=input.
- Long-running processing should be moved to a **background thread**, but the Swing view tree is confined to the event-dispatch thread. So accessing Swing objects from another thread requires using the event loop as a message-passing queue, to get back to the event-dispatch thread.



# There are many Java GUI frameworks

- **AWT** – obsolete except as a part of Swing
- **Swing** – the most widely used, by far
- **SWT** – Little used outside of Eclipse
- **JavaFX** – Billed as a replacement for Swing
- A bunch of modern (web & mobile) frameworks
  - e.g., Android



The end

May 29, 2019