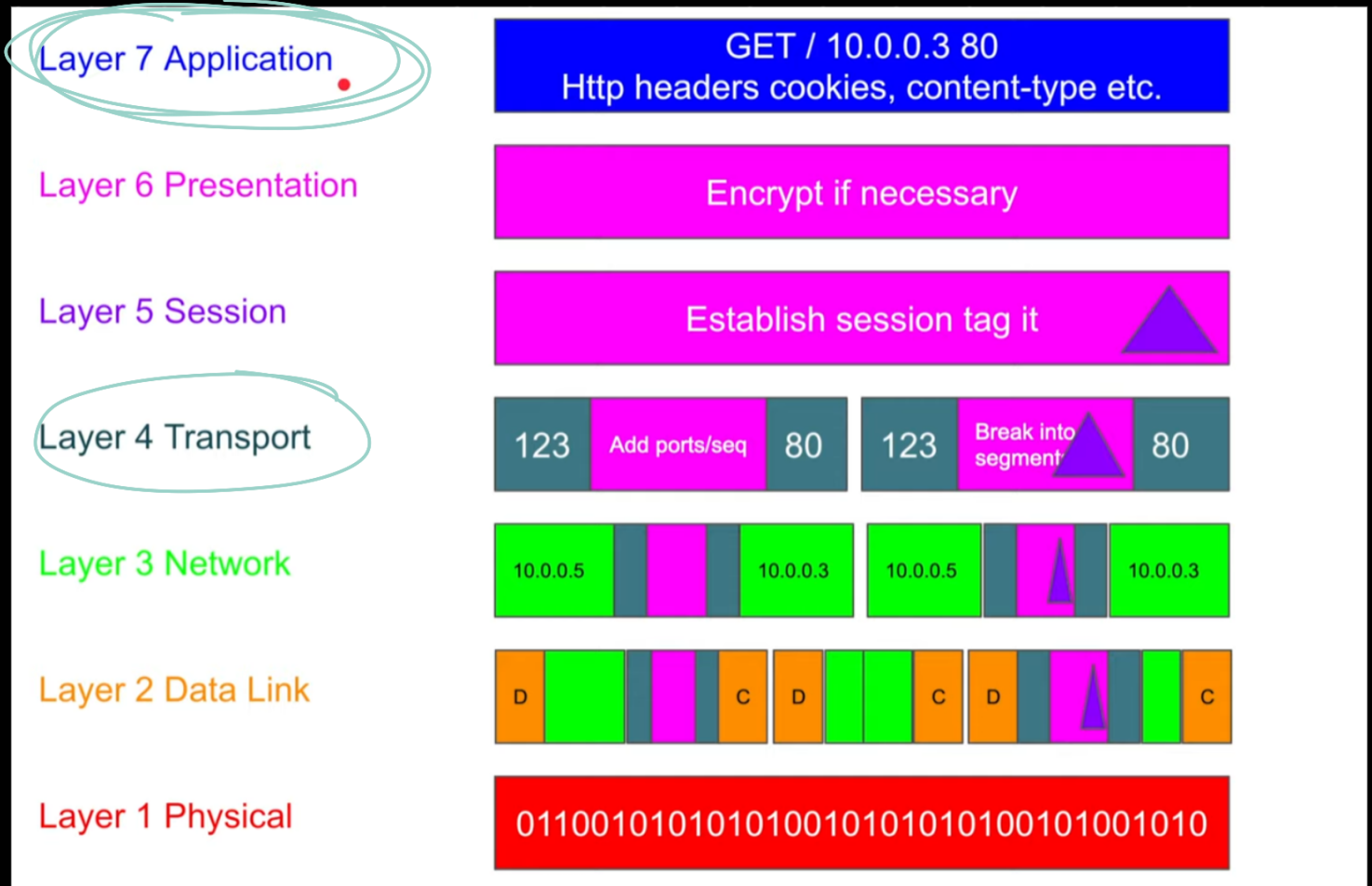


[Layer 4 v/s Layer 7 Load Balancing]



[Layer 4] → Load Balancing based on IP of servers and port

↳ It can use any algorithm of load balancing

Layer 4 Load Balancing (Transport Layer)

- How it works:
 - Operates at the transport layer, handling TCP or UDP protocols.
 - Distributes traffic based on network information like IP addresses and port numbers, without inspecting application data.
 - It forwards packets to servers without modifying them.
- Key Features:
 - **Speed:** Lower overhead as it doesn't inspect application data.
 - **Simplicity:** Straightforward implementation.
 - **Protocol Agnostic:** Works with any application that uses TCP/UDP.
- Common Use Cases:
 - Suitable for basic web servers, database traffic, or low-complexity applications.
 - High-throughput environments where application-layer inspection is unnecessary.
- Limitations:
 - Cannot make decisions based on application content.
 - Less granular control compared to Layer 7.

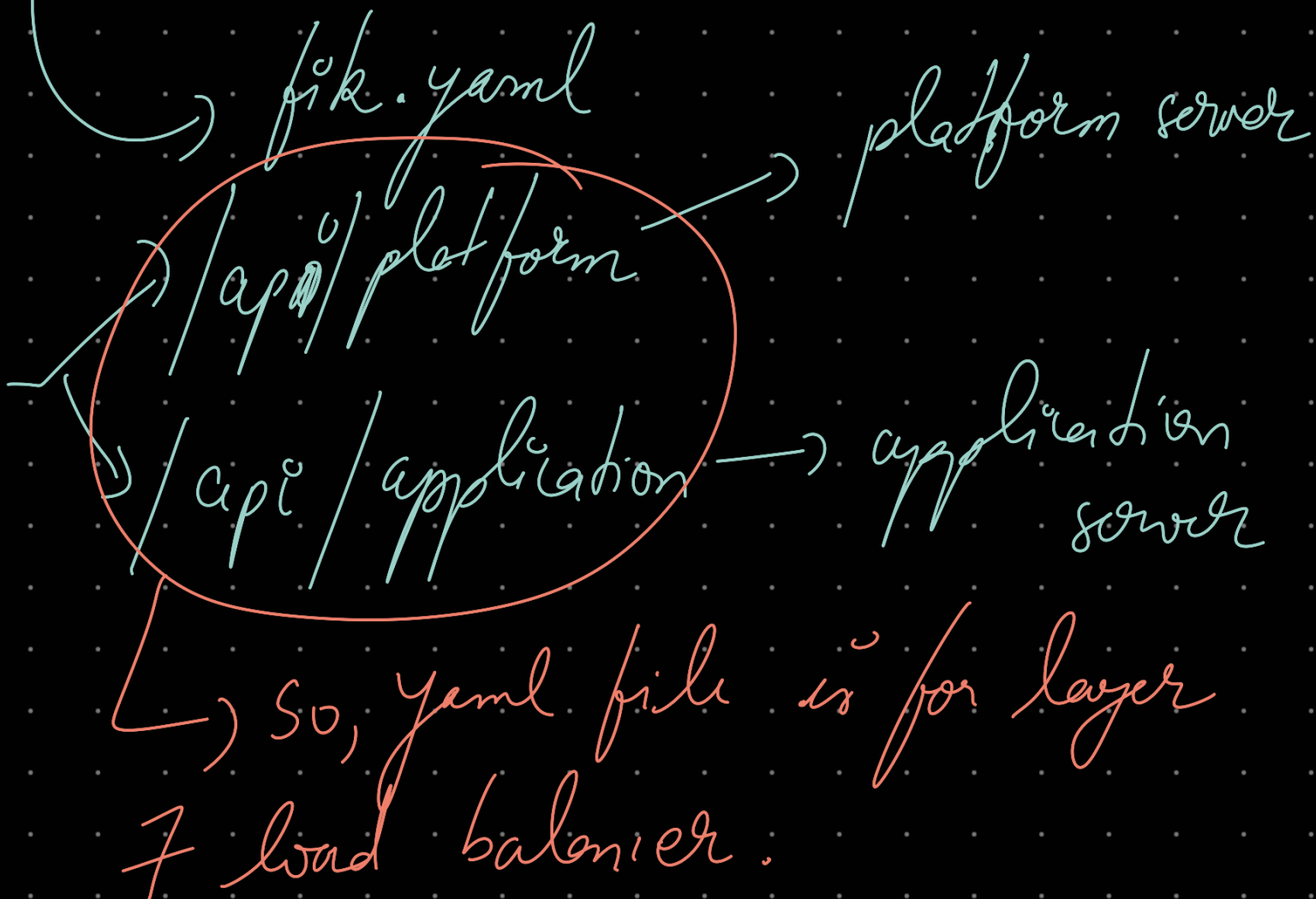
[Layer 7]

↳ Distributes on bases of application data → url, headers, cookies etc.



Layer 7 Load Balancing (Application Layer)

- How it works:
 - Operates at the application layer, analyzing HTTP, HTTPS, FTP, and other protocols.
 - Distributes traffic based on application data such as URLs, headers, cookies, or content types.
- Key Features:
 - **Content-Based Routing:** Can redirect requests based on specific application data (e.g., route images to one server, videos to another).
 - **SSL Termination:** Decrypts and inspects encrypted traffic (e.g., HTTPS).
 - **Advanced Rules:** Implements application-level rules like directing requests for `/api` to one server and `/images` to another.
- Common Use Cases:
 - Complex web applications, microservices and APIs.
 - Scenarios requiring session persistence, content-specific routing, or advanced security.
- Limitations:
 - **Performance Overhead:** More resource-intensive as it inspects and processes application data.
 - **Complexity:** Requires more sophisticated configuration and management.



→ route path is visible or it is application data.

→ Caching can be done by Layer 7