A Glance Of The Philosophy Of Mass Internet Services In Tencent

Bisonliao@tencent
Technical Director

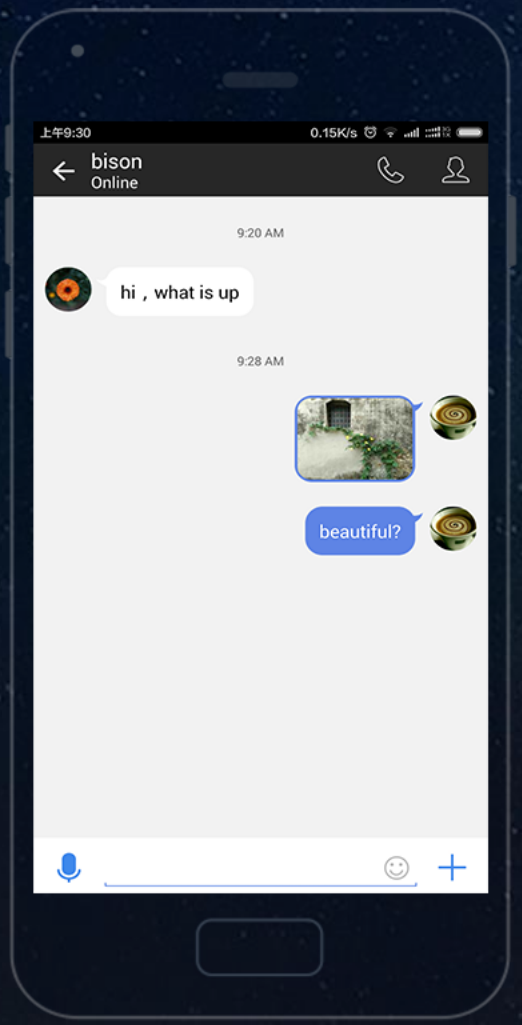
The Philosophy Of Mass Internet Services In Tencent

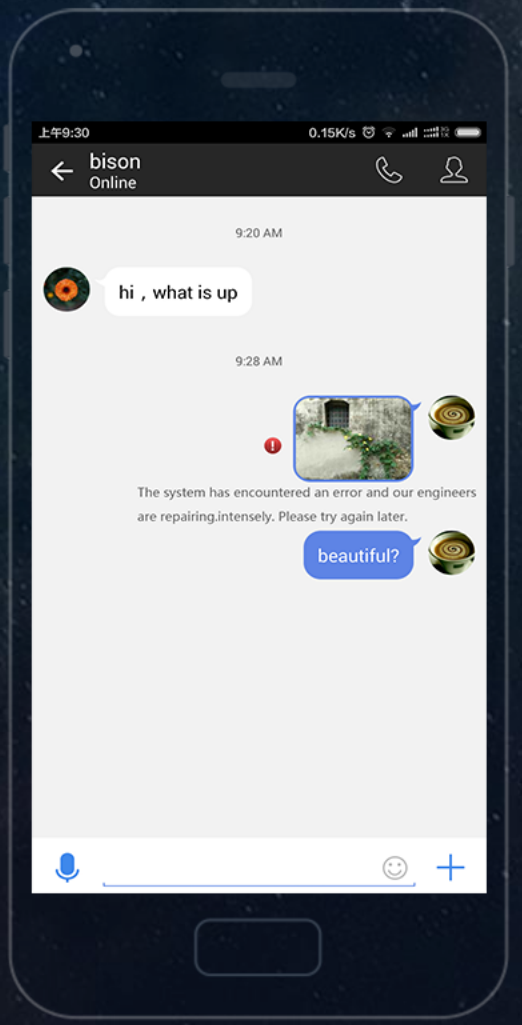
Operation by set	Overload protection	Dynamic operation
Flexible availablity	Deploy automatically	Grow in the cloud
Withstand, then optimize	Reconstruct while enjoying life	Gray-scale upgrade
Keep the system clean	Multidimensional monitoring	Imperfect services

Agenda


Flexible availability
1
2
Comfort notice
Process crash
3
4
Overload protection

What Is Flexible Availability?


上午9:30
0.15K/s
bison
Online
9:20 AM
hi，what is up
9:28 AM
beautiful?


上午9:30
0.15K/s
bison
Online
9:20 AM
hi，what is up
9:28 AM
The system has encountered an error and our engineers are repairing.intensely. Please try again later.
beautiful?

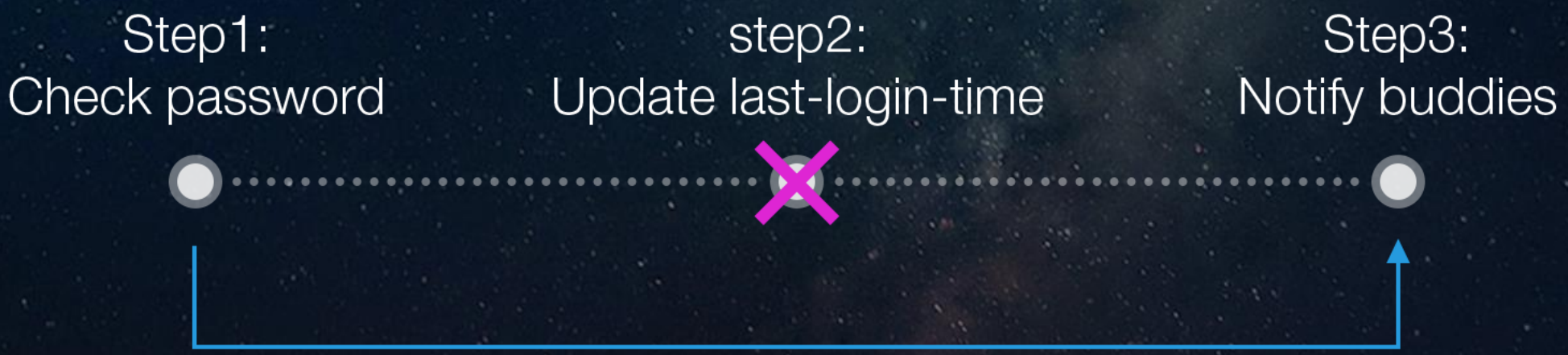
Back-end Process Of Login

Step1:
Check password

step2:
Update last-login-time

Step3:
Notify buddies

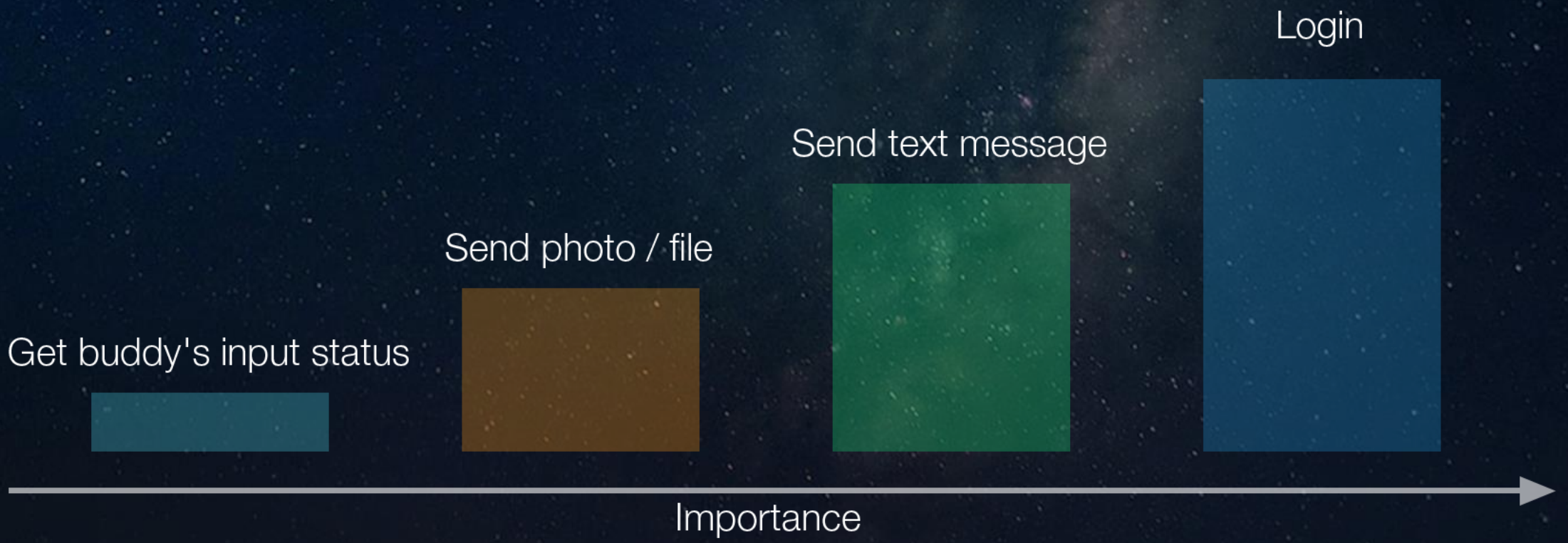
Back-end Process Of Login



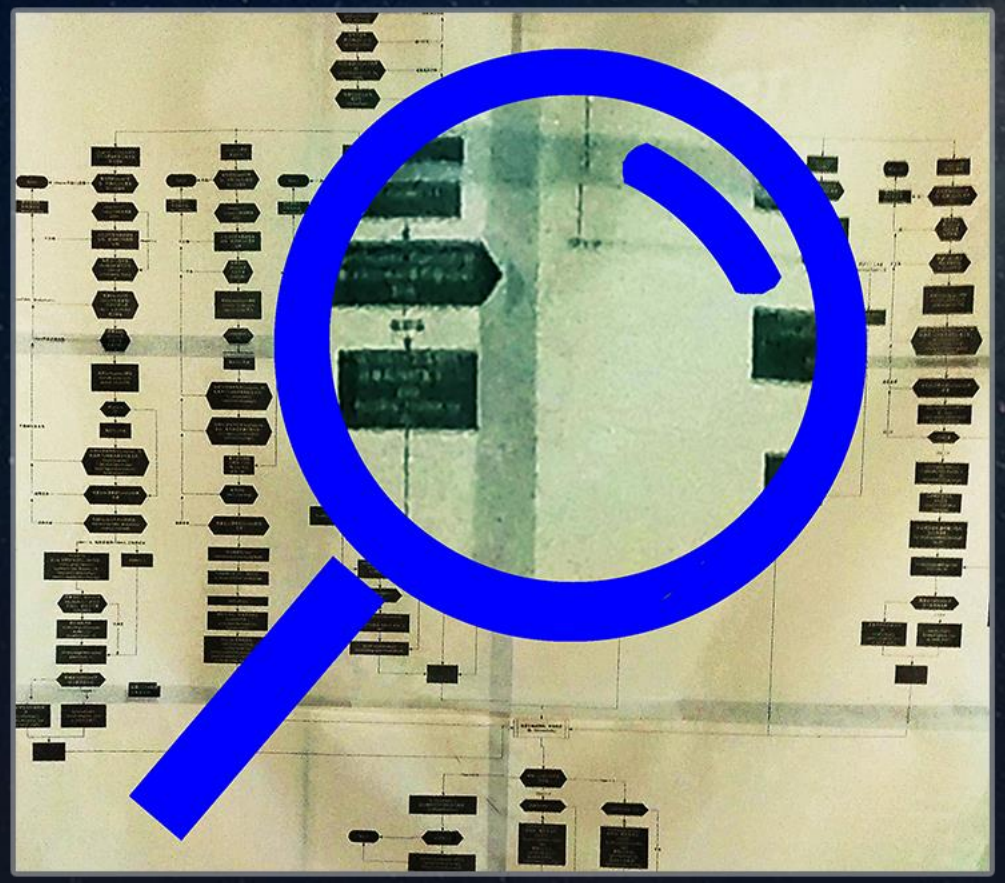

VO TOUCH

How To Achieve Flexible Availability?

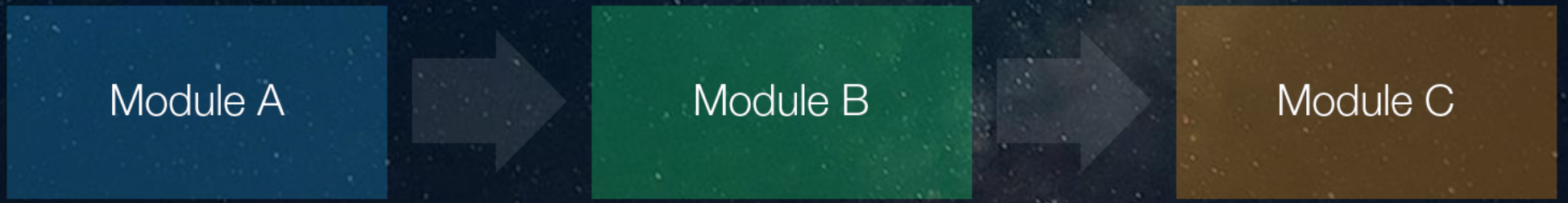
1.Grade And Decouple The Features



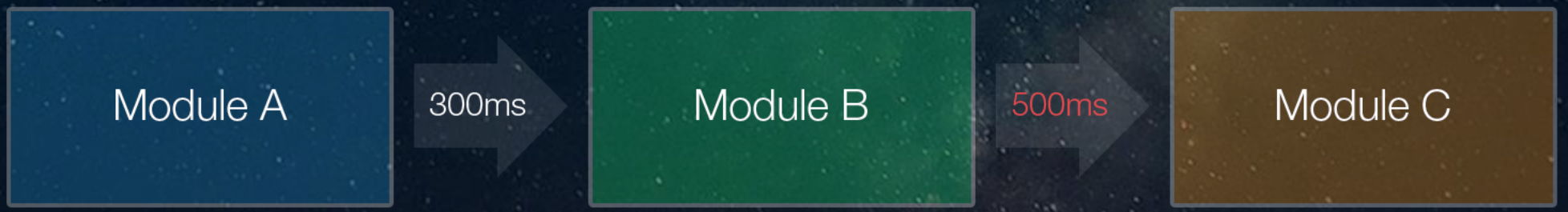
2.Deep Insight Into The Operation



3.Proper Timeout



3.Proper Timeout



Agenda


Flexible availability
1
2
Comfort notice
Process crash
3
4
Overload protection

What Is Comfort Notice?


Restaurant
Closed
Closed

Customers Confused

- Out of business?
- Passerby: “murder occurred here maybe…”
- Boss of next-door restaurant : “someone had food poisoning…”
- Customers get angry and break the windows…
- It is not yet noon, customers wait for a long time and get angrier.

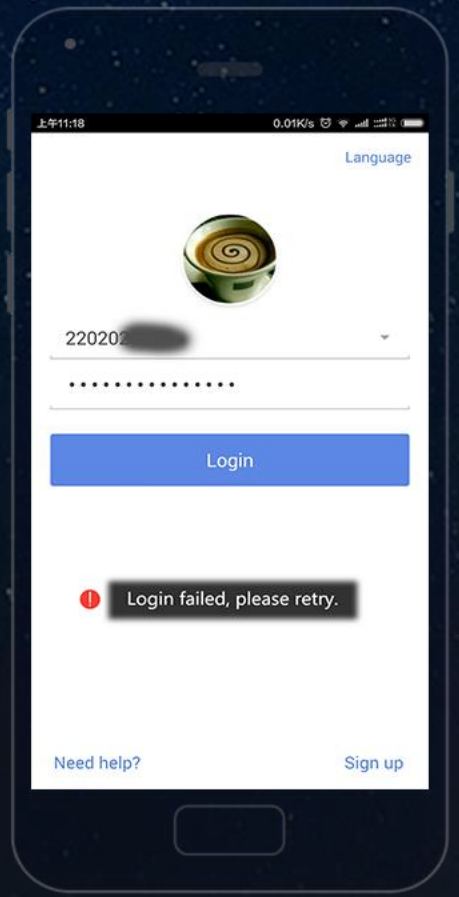
NOTICE

Due to the elopement of the wife of the boss, we will be closed for 3 days and will open on Nov. 22th.

We look forward to servicing you soon.

In Case The Bank Is Closed For 1 Day…

If The Internet Service Fails Widely



All kind of rumors :

Hacker attacked?
My password leaked?
Programmers formatted disks?
Data center power down?
…

Alternatives?

Comfort notice is NOT:

- Official micro blogs,
- Customer service center,
- Press conference

The Effect Of Comfort Notice

- Avoid customers confusion
- Avoid smears by rivals
- Relieve maintenance engineers' pressure
- Avoid back-end system overload by users' retry

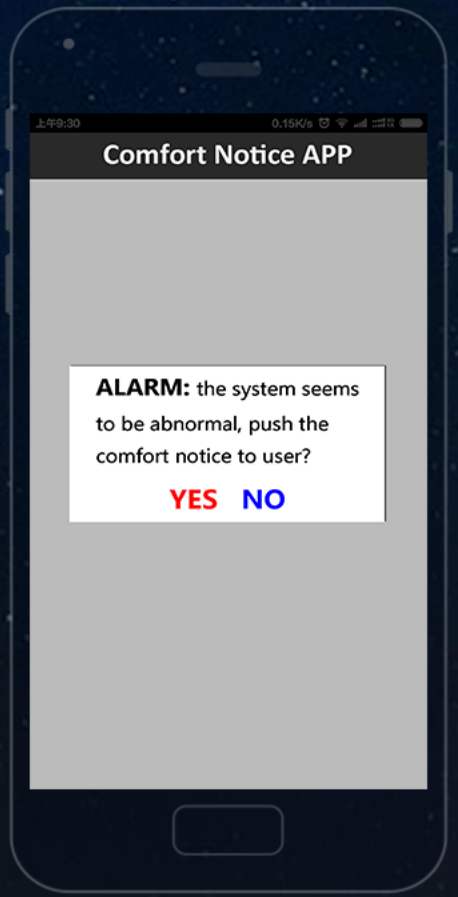
How To Operate Comfort Notice?

1. Independent

2. Automatic, Self-executing

3. Robust And Simple

Example



Login server

Step 1: login,
sometimes failed or timeout

User

Step 2: need to show notice?

Step 5: push notice to users

Maintainance engineers

Step 3: need to push notice?

Step 4: config notice to push

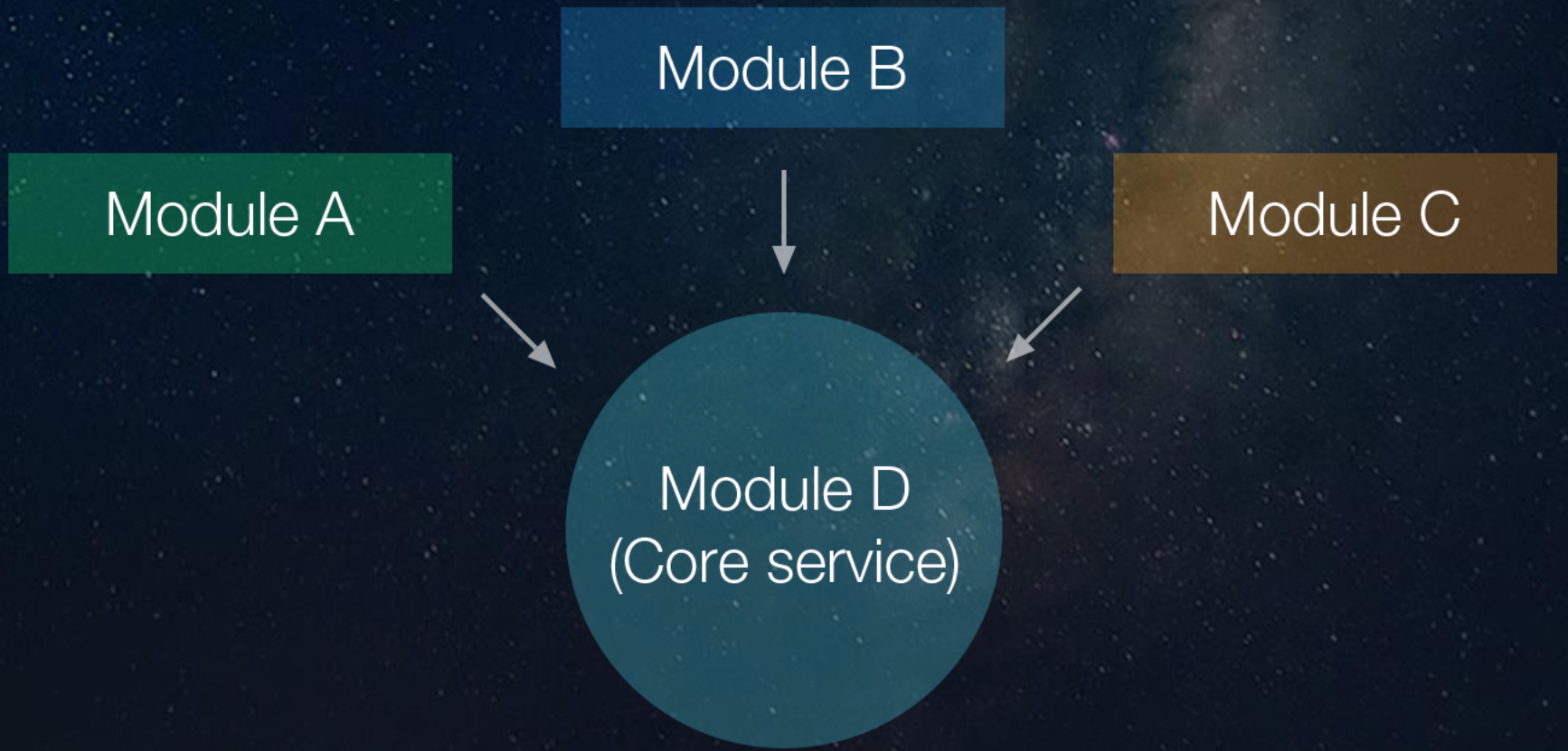
Comfort notice server

Agenda


Flexible availability
1
2
Comfort notice
Process crash
3
4
Overload protection

Case Study (I)

Case Study (II)

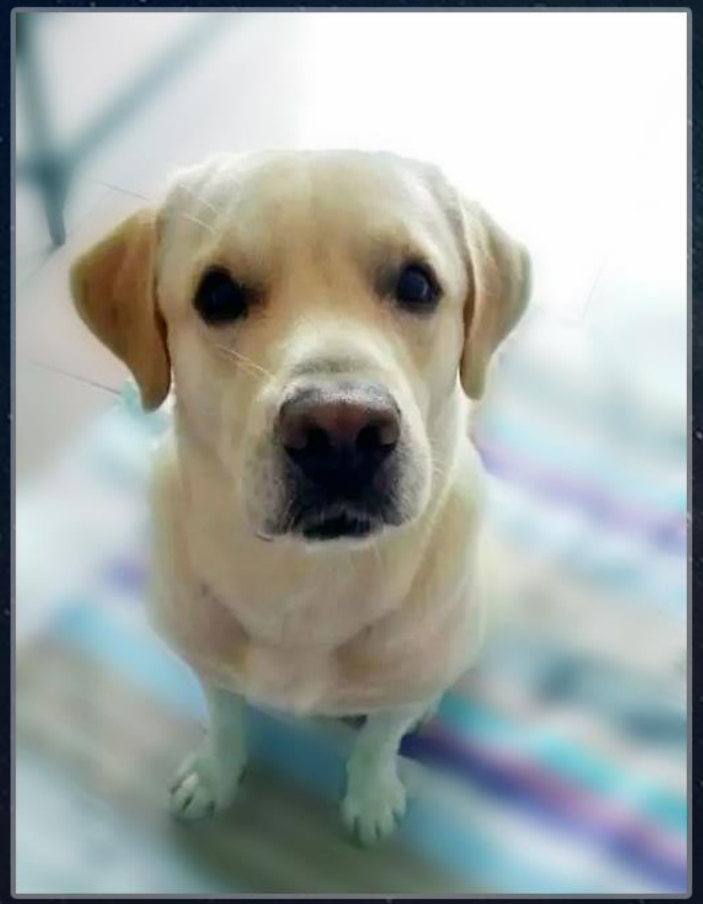


Some Guidelines...

- Defensive programming
- Code reuse (IDL like protocolbuffer)
- Fully test
- Communication between teams

But NOT Enough

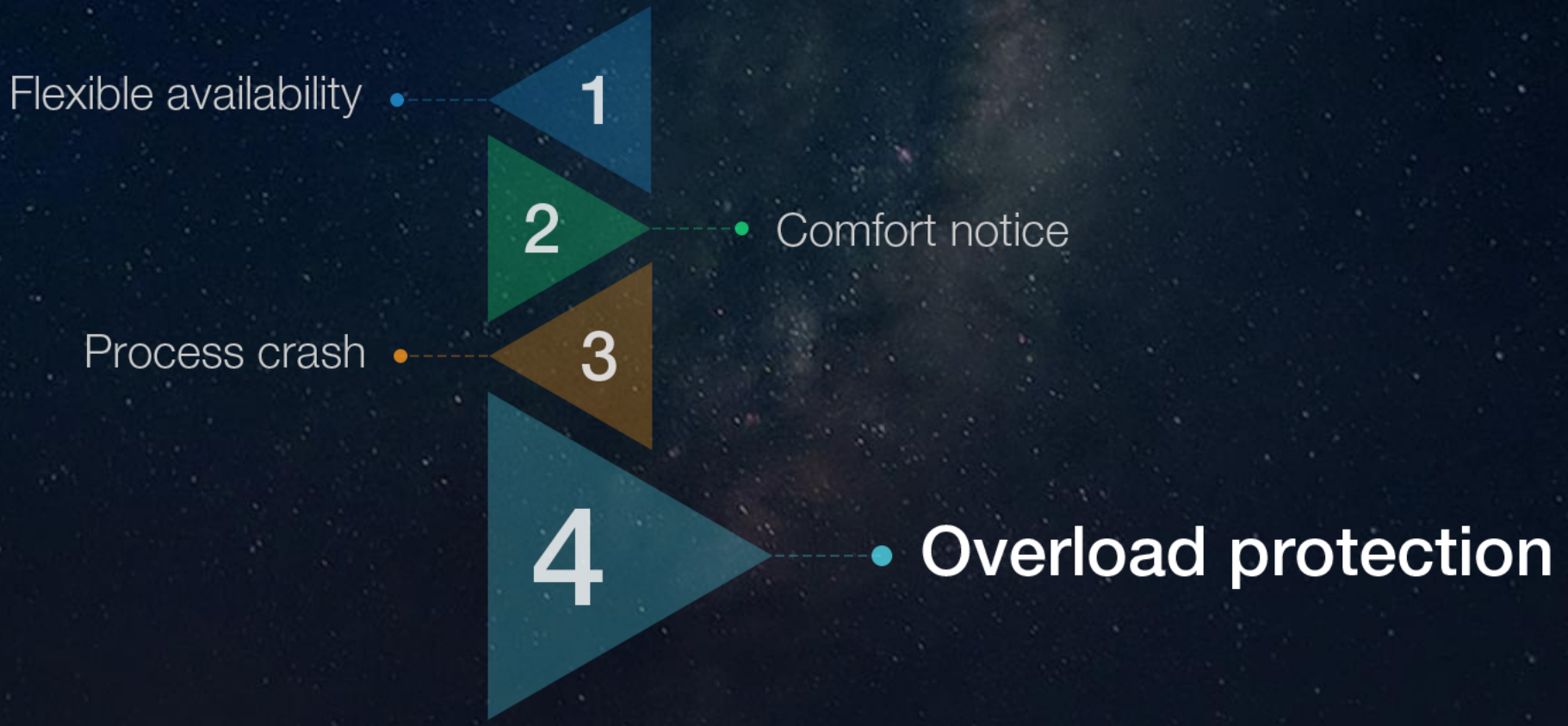
1. Process Monitoring, Restart In Milliseconds



2. Set Model For Public Service

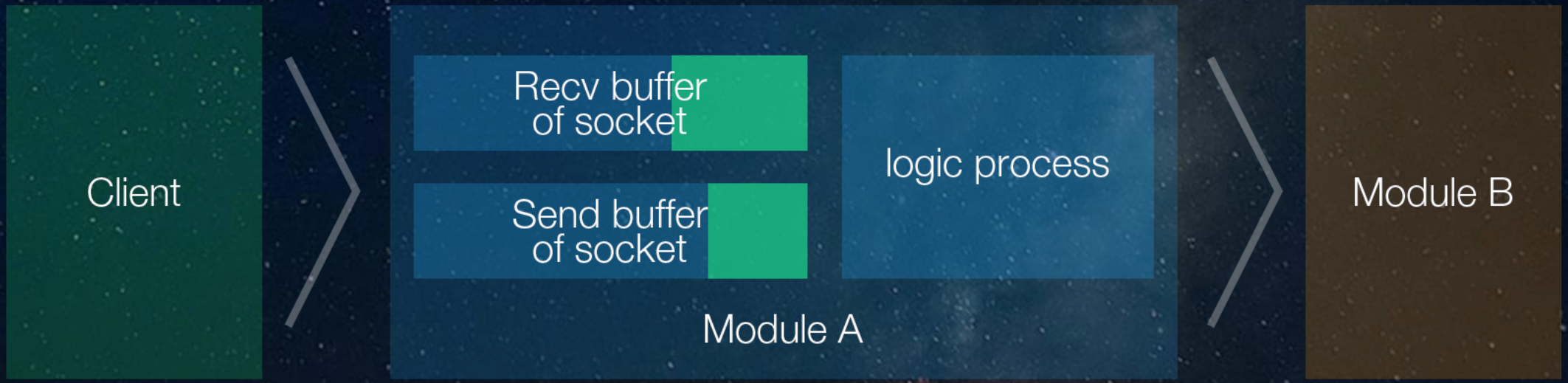
Just like there is a fire (crash) resisting curtain

Agenda


Flexible availability
1
2
Comfort notice
Process crash
3
4
Overload protection

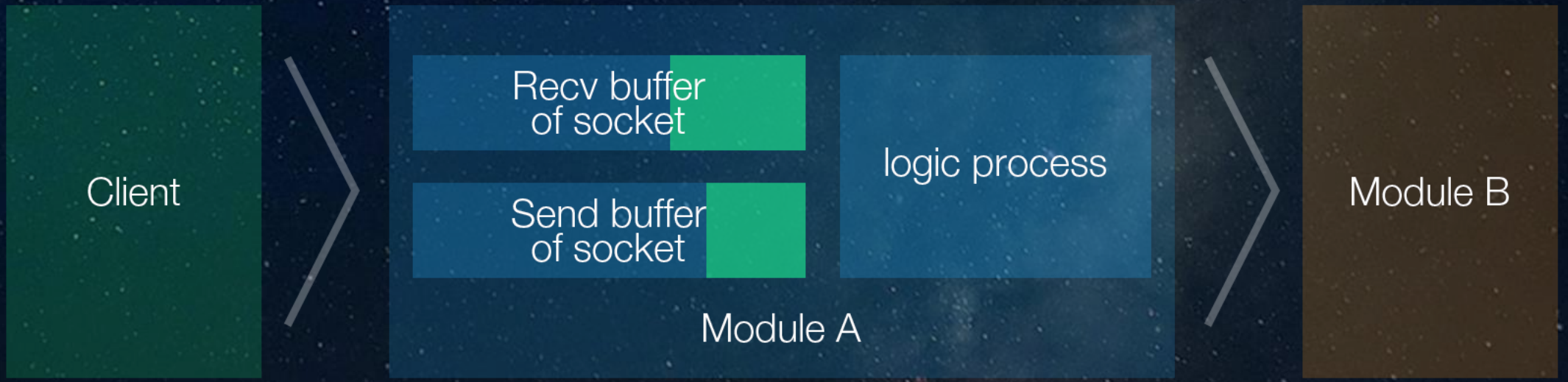
What Is Overload?

Case Study



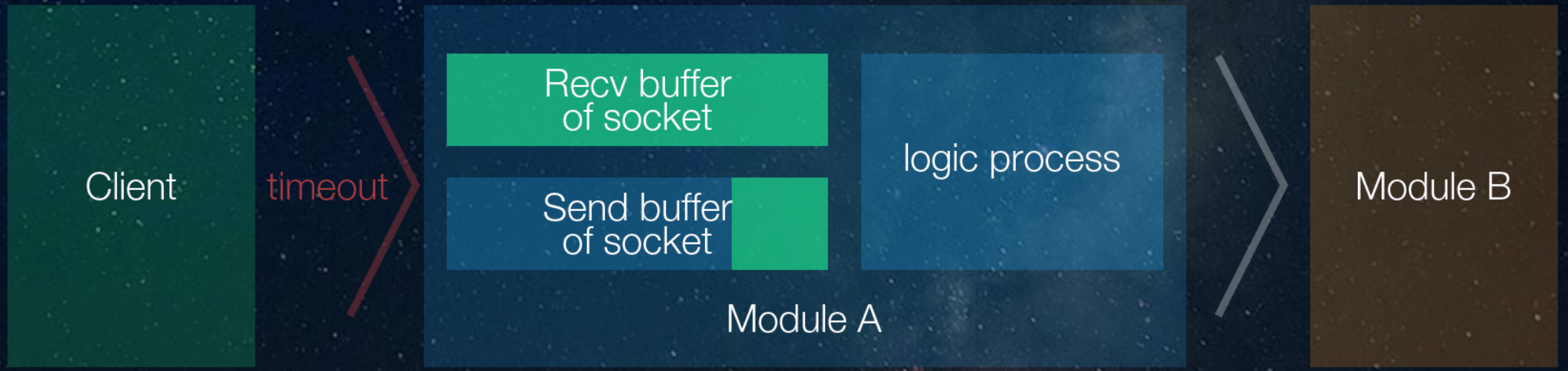
The business logic of module A is as follows:
Step 1: Get the user requests from receive Buffer.
Step 2: Perform the local logic processing.
Step 3: Send requests to Module B and wait for the response.
Step 4: Process the response from Module B.
Step 5: Send response to clients ,and goto step 1 for next request.

Case Study



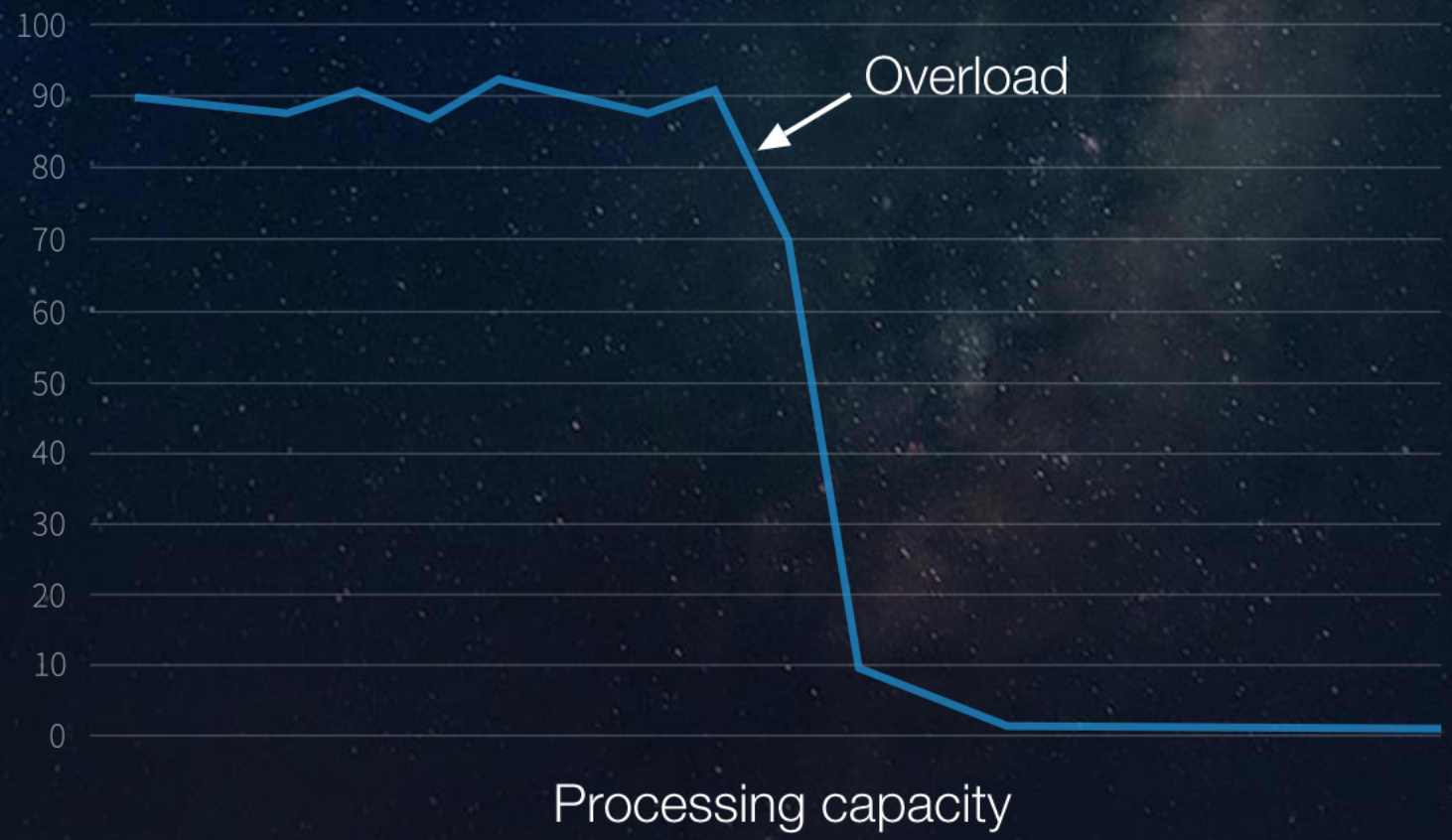
The requests can be handled in 20 milliseconds, module A can handle up to 50 requests per second, which is sufficient for peak client requests as much as 30.

When Buffer Becomes Full...



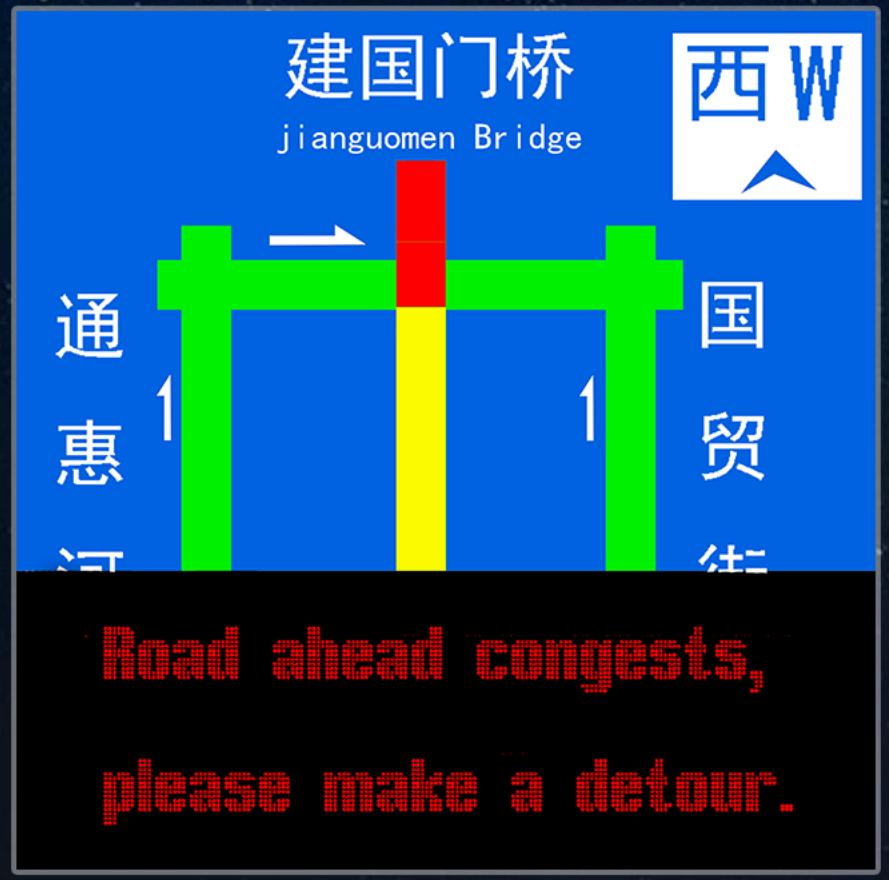
1. One day, module B releases a new feature and the average request processing time increases from 20 milliseconds to 50 milliseconds.
2. Users who get a failed notice would usually retry, buffer is kept full all the time.
3. when a client request is queued in the receive buffer, it has to wait 5 seconds until module A can fetch it.

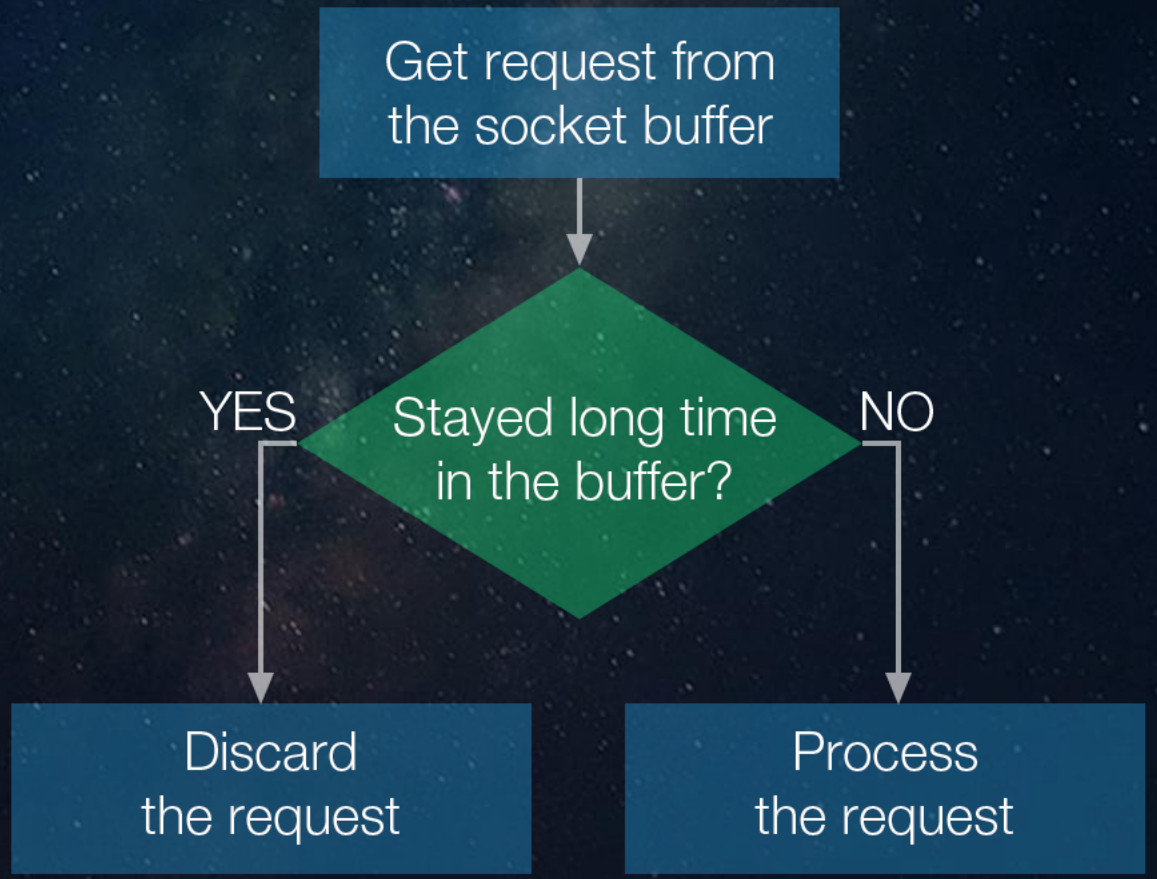
Effective Capacity Drops To 0



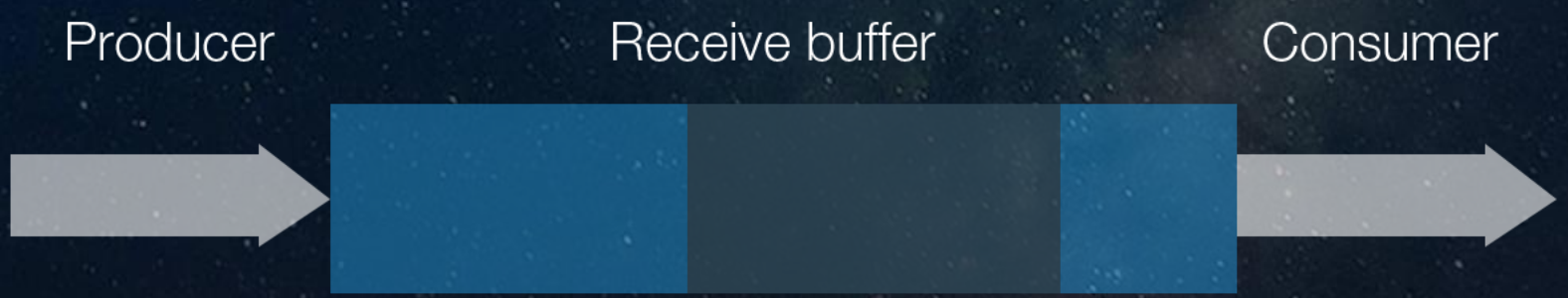
How To Achieve Overload Protection?

1. Refuse Early

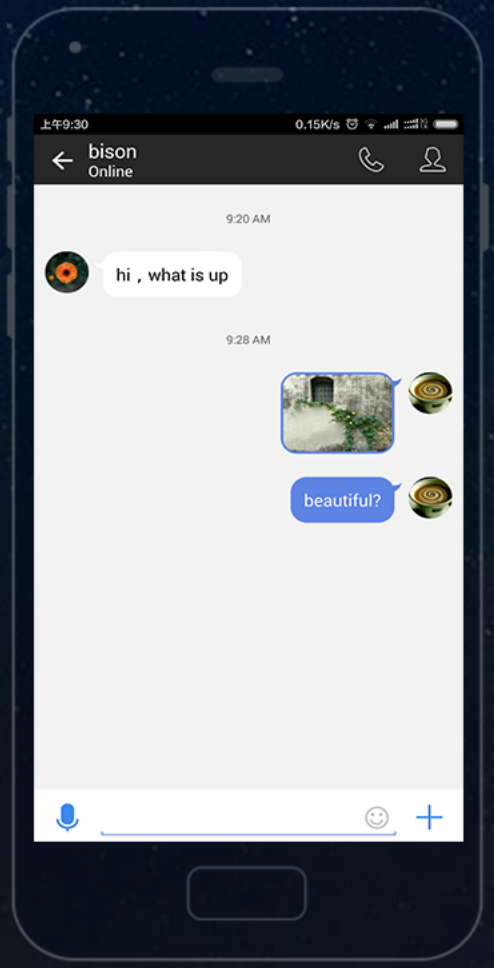




2. Proper Recv Buffer Size

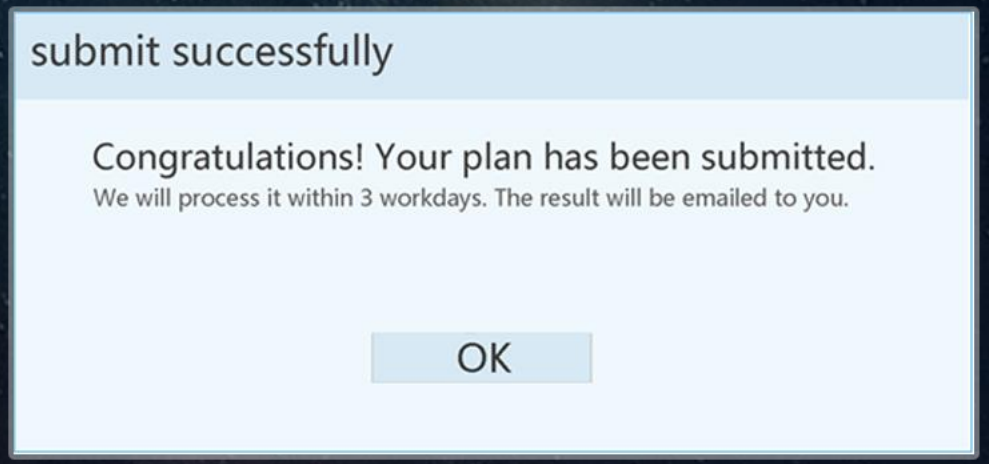


3. Retry With Limits




```
iRetCode = AccessModuleB();
if (iRetCode == ERR_TIMEOUT)
{
    if (SuccessRateInLast5Min > 0.98)
    {
        // retry
        AccessModuleB();
    }
}
```

4. Smart Flow Design



The End