全用户态服务开发套件F-Stack

演讲者：姜凤波

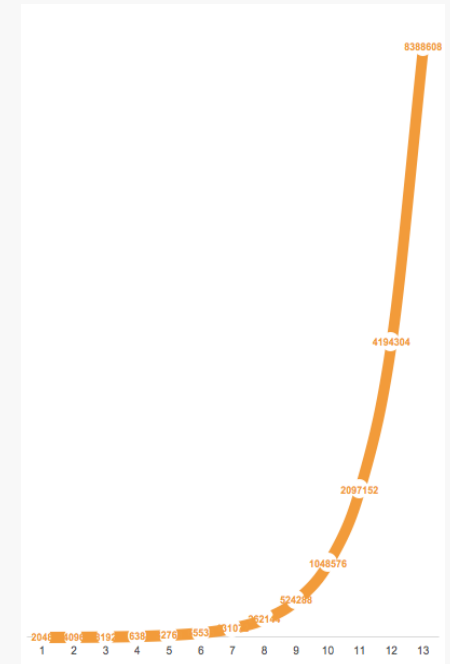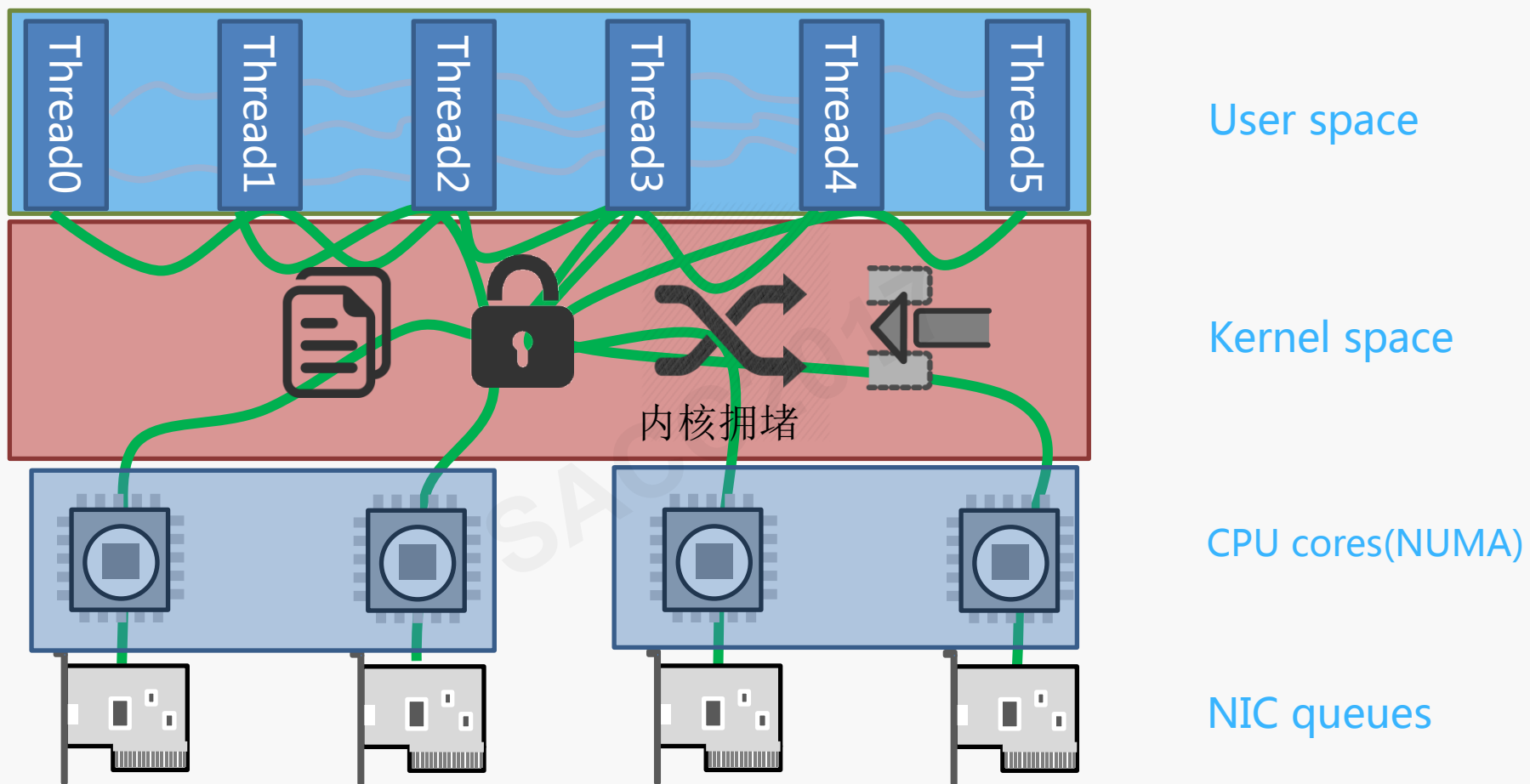# Explosive growth of data services

▶ Challenges

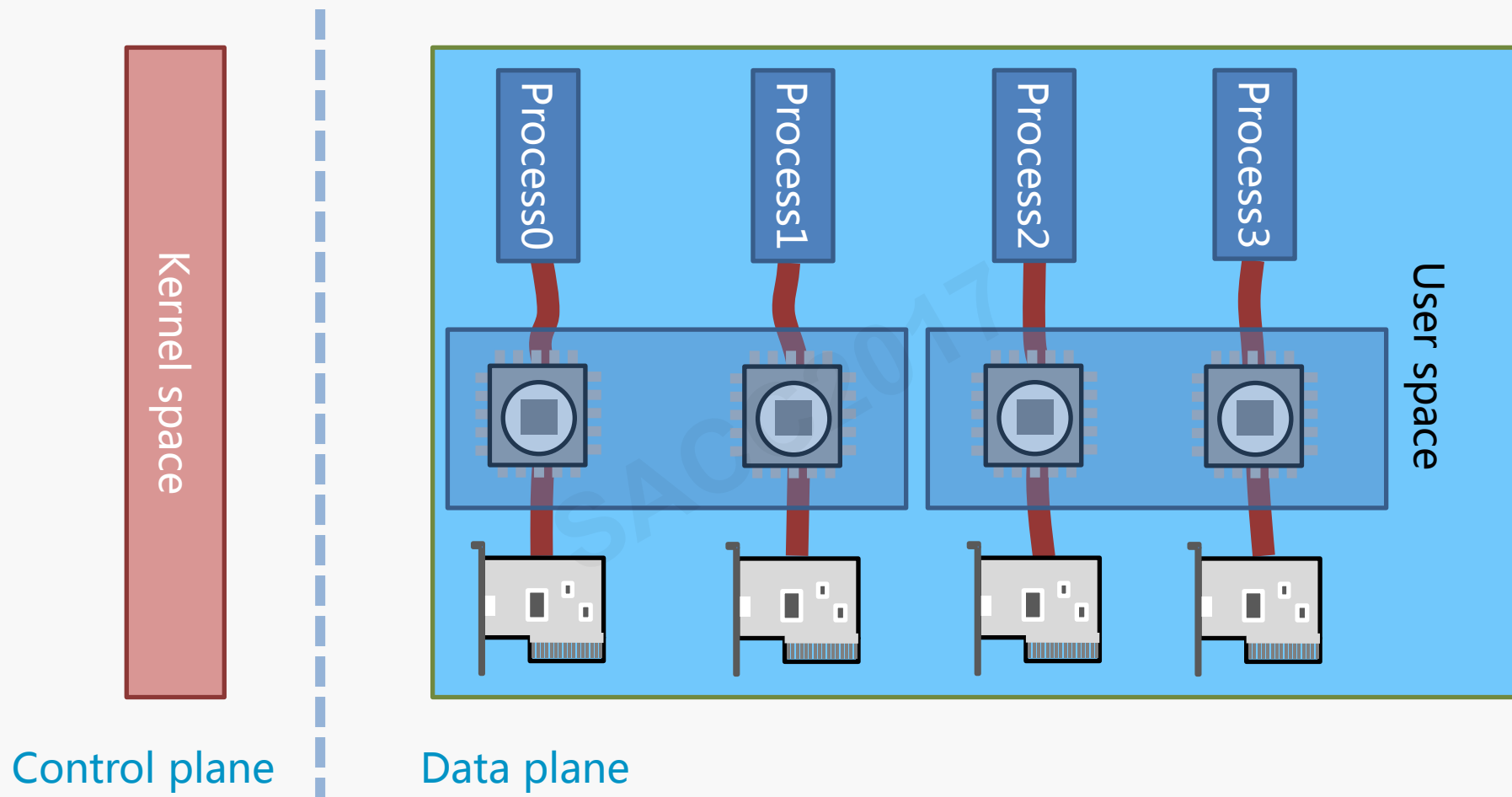   ▶ service traffic:CDN,live streaming(RTMP),etc.

    ▶ DDoS, CC

▶ Solution

    ▶ NIC: 25GbE→40GbE→100GbE

    ▶ CPU: 24Cores→48Cores→56Cores

    ▶ L2/L3: Kernel bapass(DPDK)

    ▶ L4/L7: ?

The problem - Kernel

User space

Thread0 Thread1 Thread2 Thread3 Thread4 Thread5

内核拥堵

Kernel space

CPU cores(NUMA)
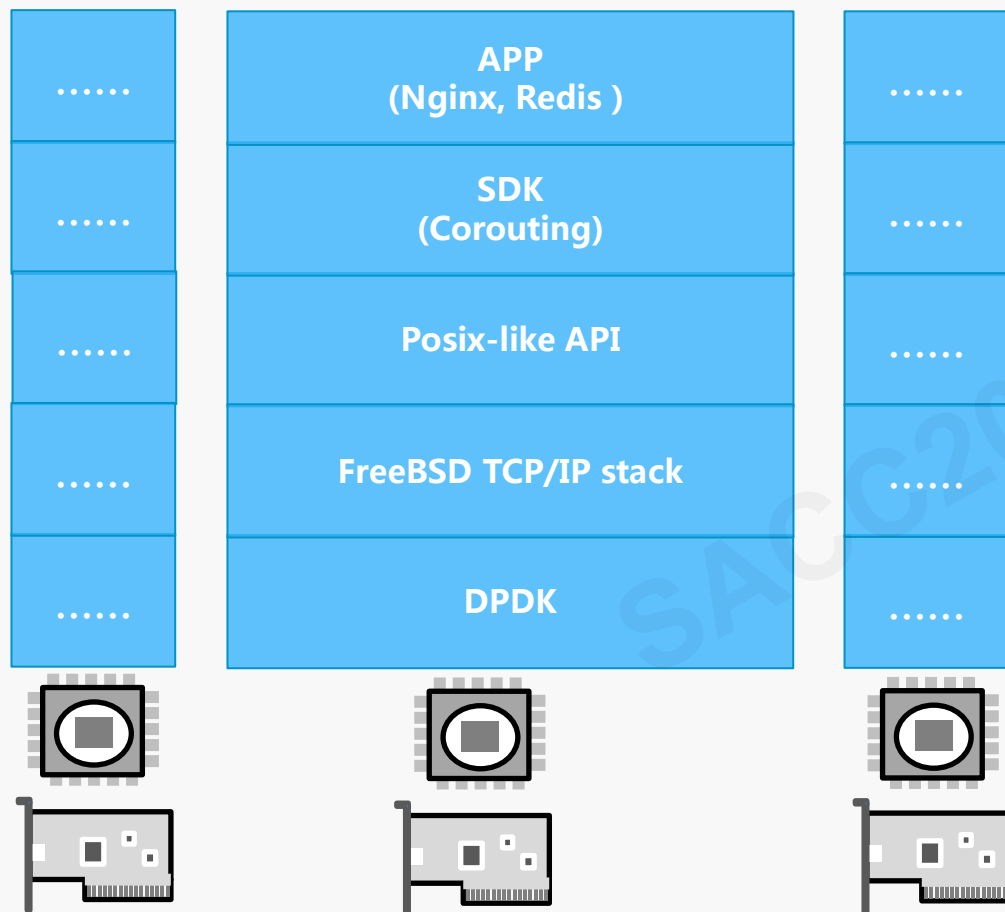
NIC queues

Kernel Bypass

User space TCP/IP stack

▶ Seastar's native stack

　　▶ problems in WAN

　　▶ incompatible with existing programs

▶ mTCP's user-level stack

　　▶ tcp only.

　　▶ too simple to be a production ready stack

VLAN/VxLAN/Tunnel/Bonding/Network Tools are not supported

What does F-Stack do？

APP
(nginx, redis, memcache)

SDK
(corouting)

Posix API

FreeBSD TCP/IP stack

DPDK

F-Stack
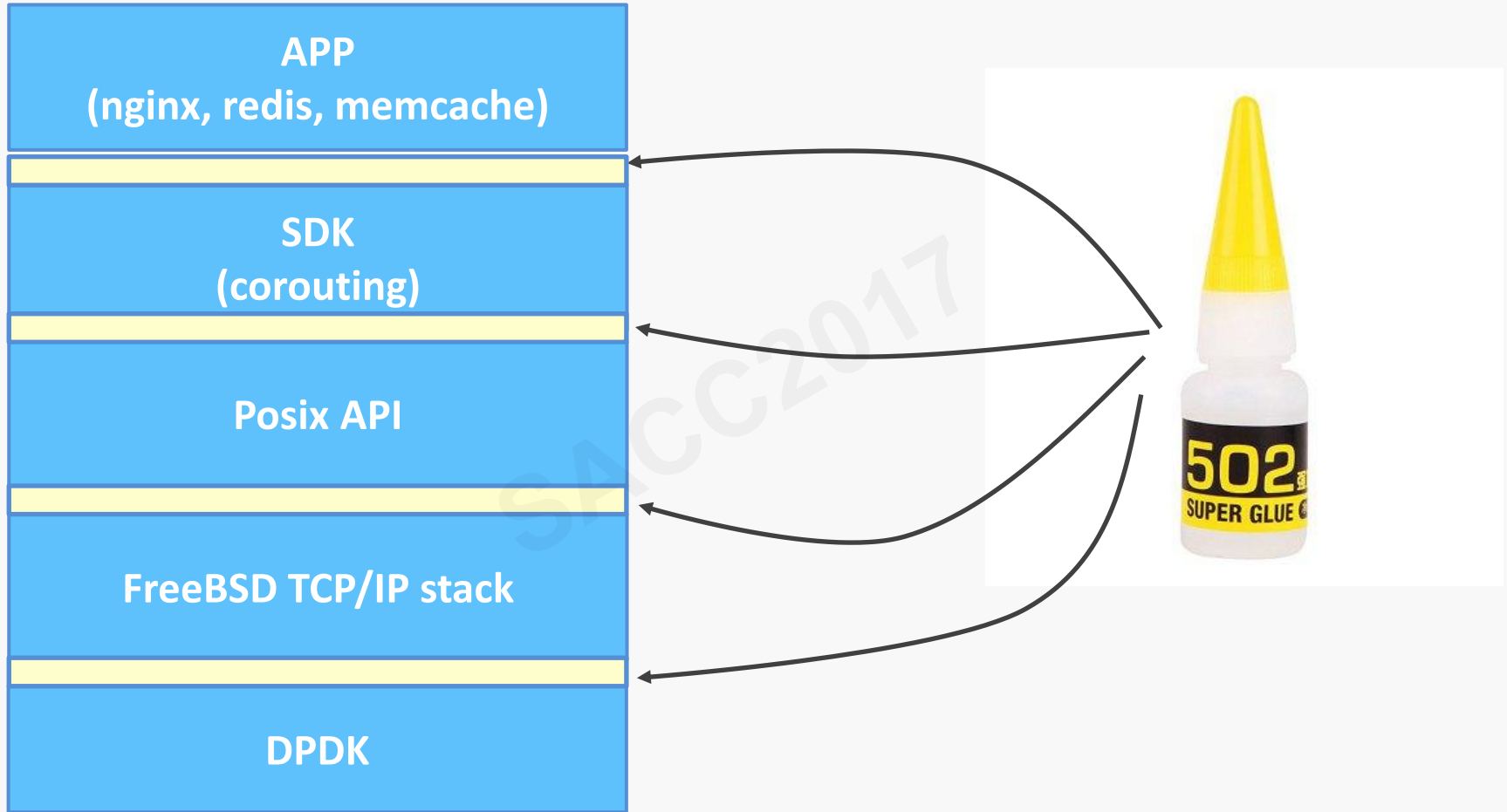
	APP (Nginx, Redis)	
......	SDK (Corouting)	......
......	Posix-like API	......
......	FreeBSD TCP/IP stack	......
......	DPDK	......

1、 Full user-space

- ▶ No content switch
- ▶ Zero copy
- ▶ No hirqs and sirqs

2、 Shared-nothing architectur
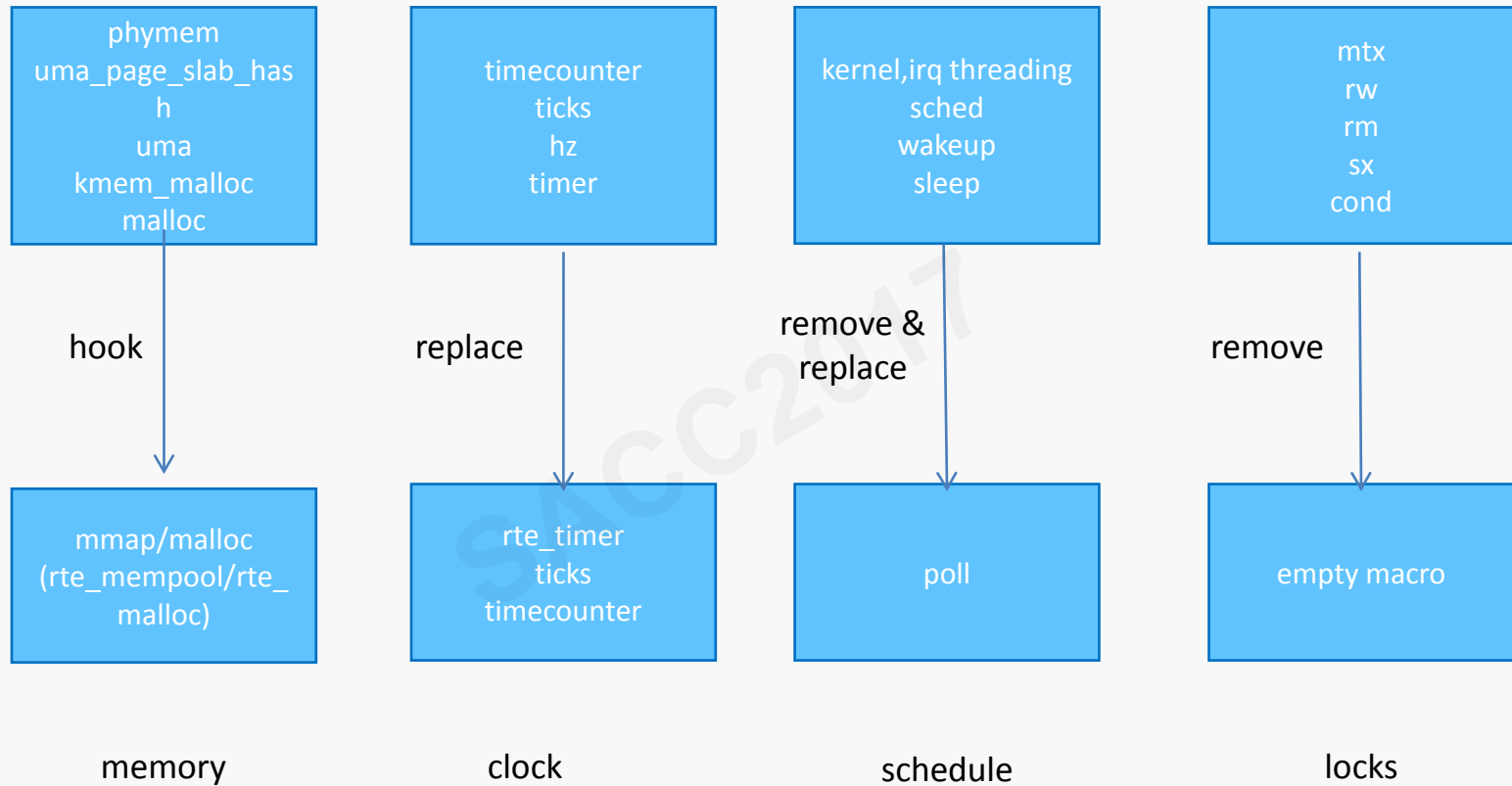
- ▶ Linear scalability
- ▶ No schedule
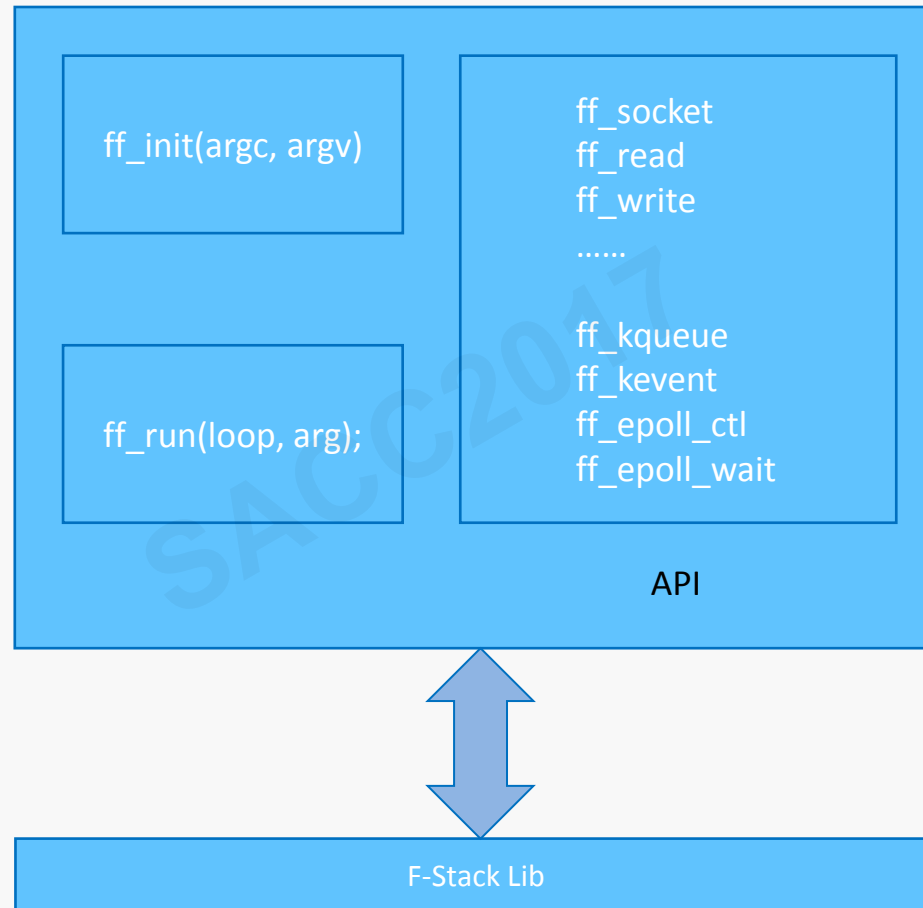- ▶ No lock
- ▶ No cache locality miss

Why FreeBSD?

▶ Why full TCP/IP stack？

 ▶ Advantageous functional, production ready stack, IGMP/IPv6/VLAN/VxLAN/Tunnel/Bonding/Net Tools

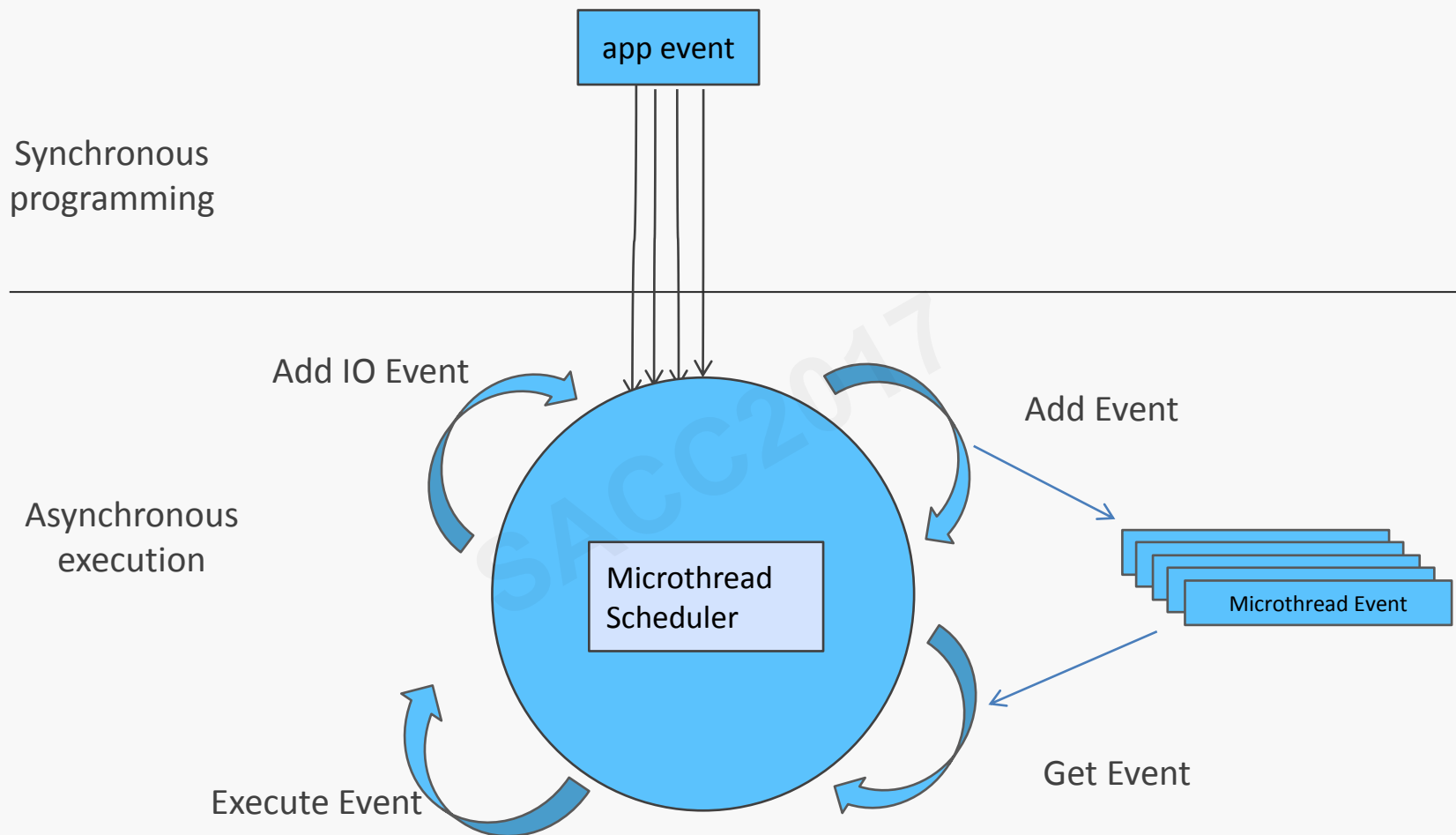 ▶ Stable.

▶ Why not Linux？

 ▶ complicated logic.

 ▶ GPL

User space FreeBSD TCP/IP stack

phymem uma_page_slab_hash uma kmem_malloc malloc	timecounter ticks hz timer	kernel,irq threading sched wakeup sleep	mtx rw rm sx cond
hook	replace	remove & replace	remove
mmap/malloc (rte_mempool/rte_ malloc)	rte_timer ticks timecounter	poll	empty macro
memory	clock	schedule	locks

Posix-like API

Coroutine

app event

Synchronous programming

Asynchronous execution

Add IO Event

Add Event

Microthread Scheduler

Microthread Event

Execute Event

Get Event

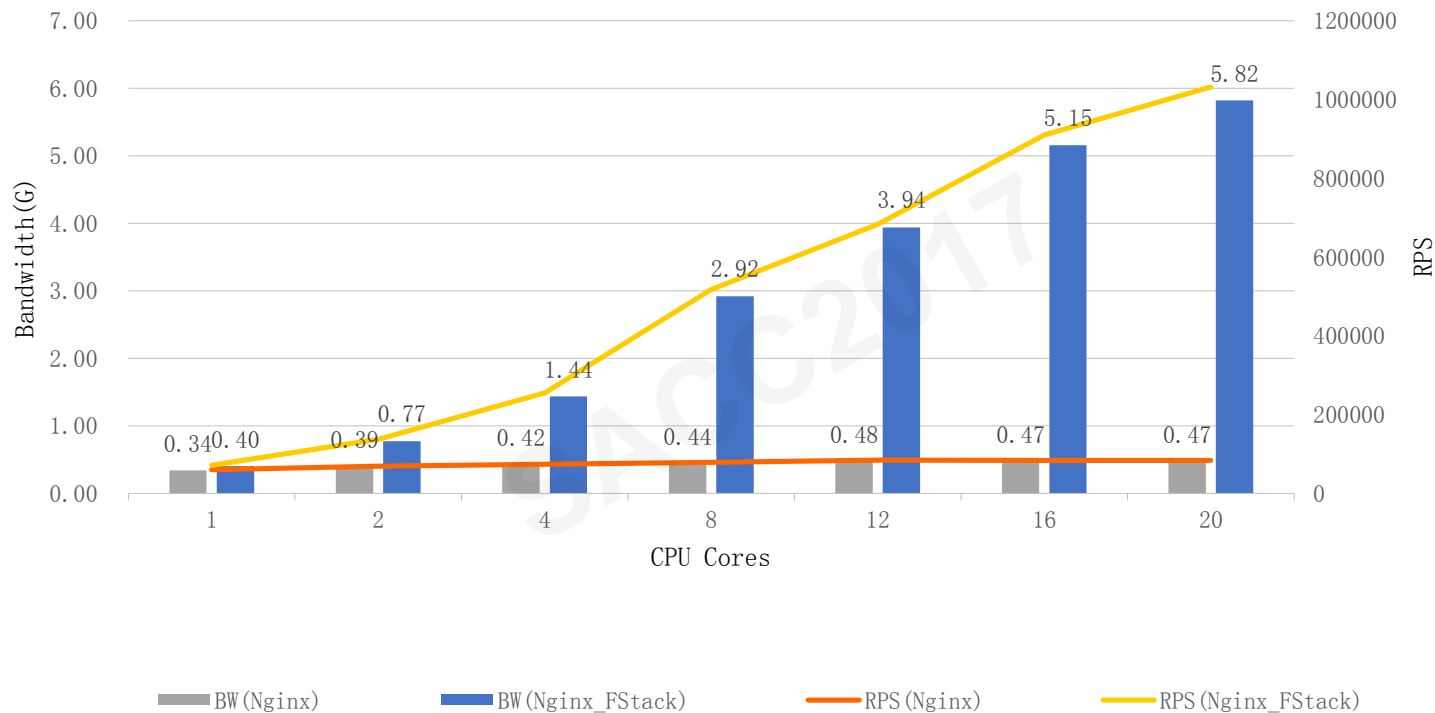# Tools

- ▶ sysctl

- ▶ ifconfig

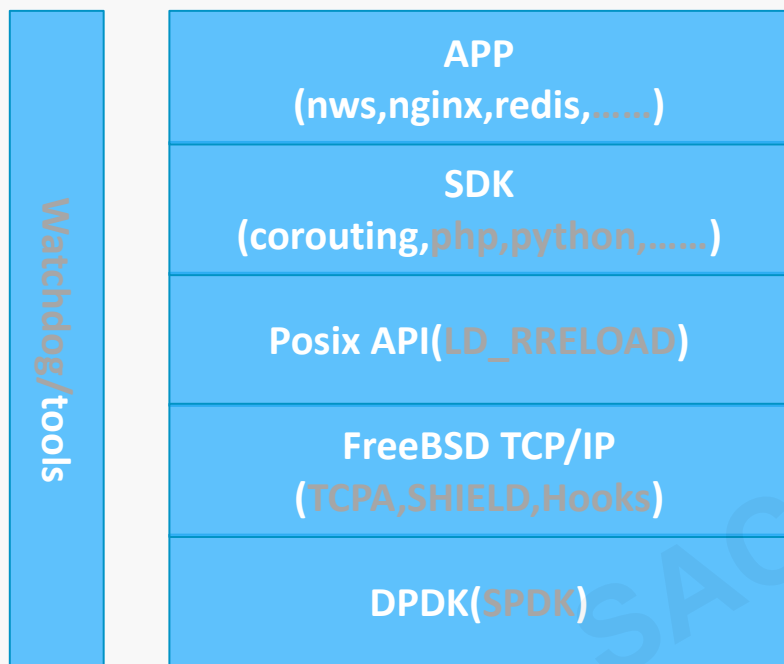- ▶ route

- ▶ netstat

- ▶ top

- ▶ etc..

Nginx_F-Stack

cache body RPS test(No Keep Alive)

F-Stack in Tencent Cloud

▶ 2013 UDP DNS server based on DPDK

▶ 2014 TCP(user-space TCP/IP stack) DNS server

▶ 2015 L4/L7 gateway application

▶ 2016 Used on HttpDNS,CDN,COS(Cloud Object Storage)

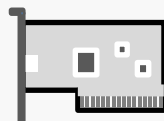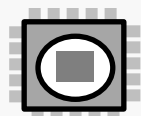▶ 2017 Port FreeBSD network stack to F-Stack and open source

Roadmap

Watchdog/tools	**APP** **(nws,nginx,redis,......)**
	SDK **(corouting,php,python,......)**
	Posix API(LD_RRELOAD)
	FreeBSD TCP/IP **(TCPA,SHIELD,Hooks)**
	DPDK(SPDK)

https://github.com/f-stack

http://www.f-stack.org

F-Stack WeChat subscription

THANKS