Inside Erlang VM

Yu Feng
mryufeng@gmail.com
2010/04/26
www.yufeng.info

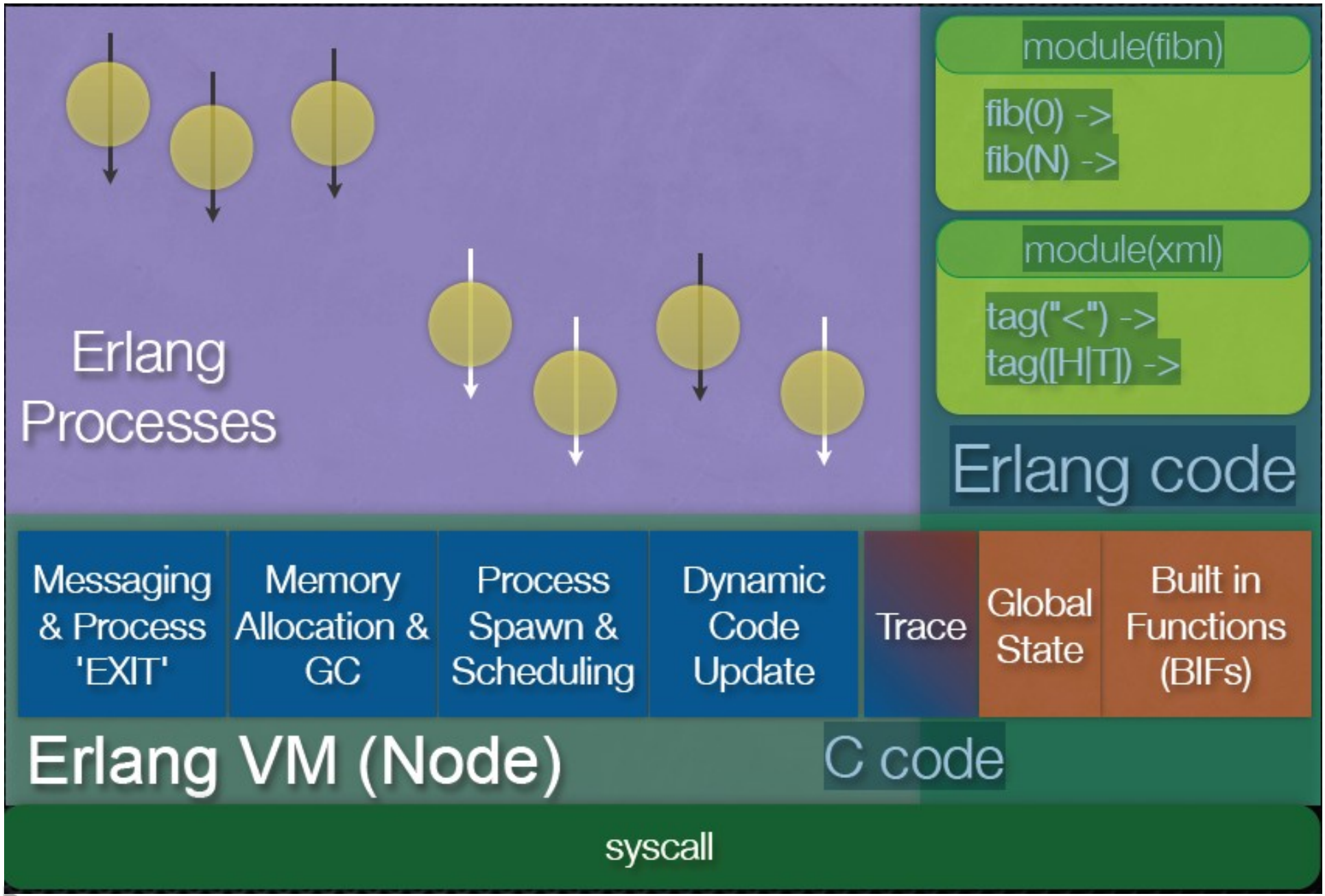
目的

通过对Erlang VM的内部部件的解刨，　能够对VM的运作原理有个深度的把握和了解，对系统的设计和实现有指导作用.

包含的内容:

- Erlang是什么
- ERTS的优势
- 高级网络程序几个要素,ERTS如何实现
- Erlang VM的特点
- Erlang集群的设施
- Erlang自省机制
- 代码的热升级
- 内置强大的数据库
- 稳定和移植性

Erlang是什么?FP语言还是平台



ERTS的核心优势

- 高性能
- 多核心SMP的支持
- FP语言支持，自动GC
- 透明分布的支持
- 轻量进程的支持
- 完善的系统信息
- 商业产品上经过时间的验证成熟

典型网络服务器几大要素

- CPU
- IO
- 消息传递
- 定时器
- 业务抽象

ERTS就是个网络程序的框架

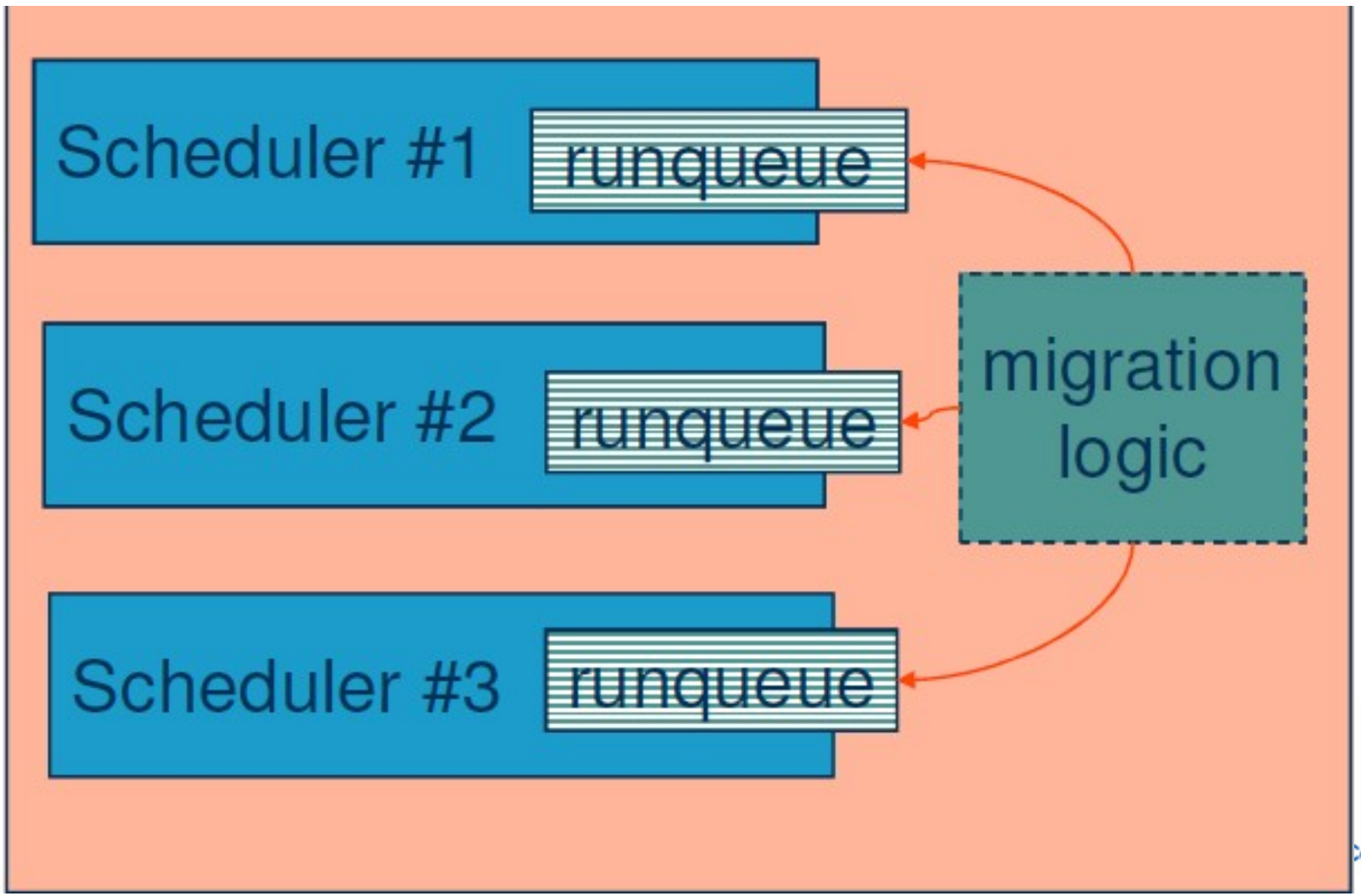
谁说的?我说的

CPU

- 多核心，多调度器
  - running on full load or not
- 抢占式调度
- 调度器公平调度
- 保持少量的CPU忙
- CPU亲缘性/进程绑定

CPU续

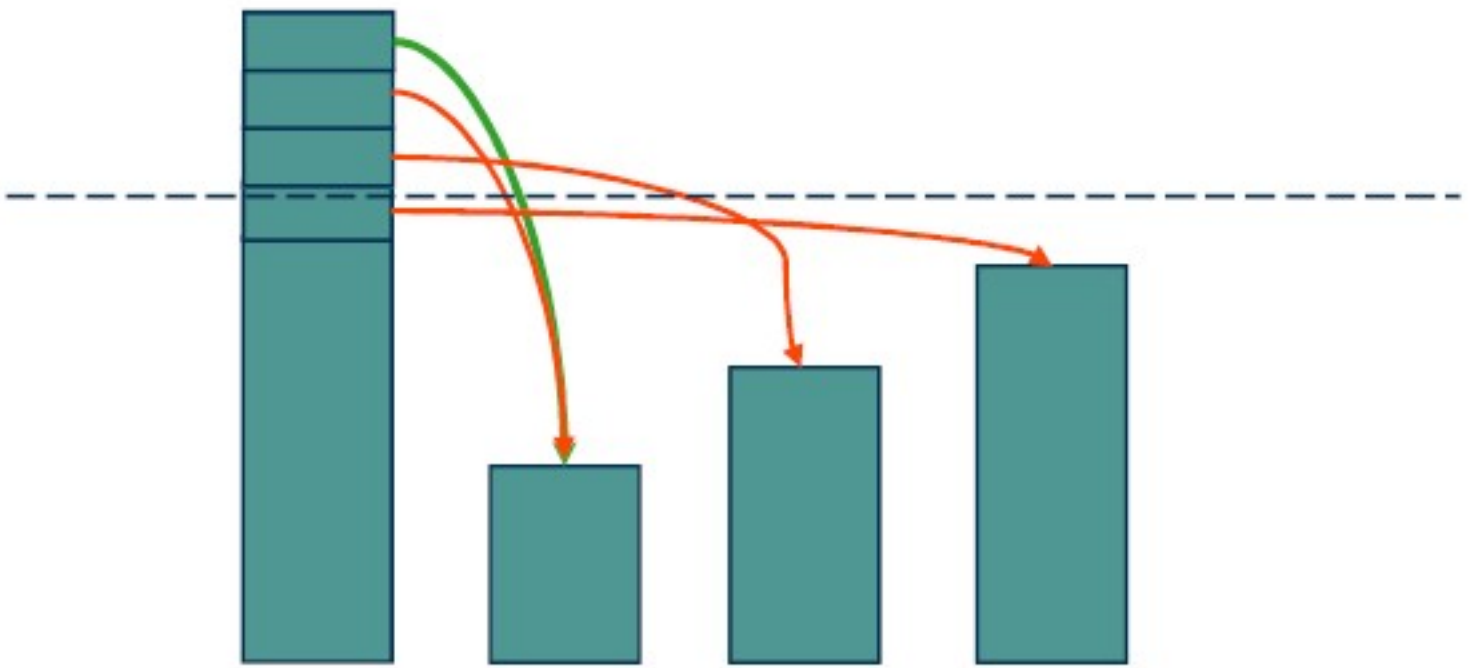
- 每个核心一个调度器
- 每个调度器一个调度队列



ERLANG
Concurrent Distributed FP Language

cpu续

CPU负载均衡和迁移



ERLANG

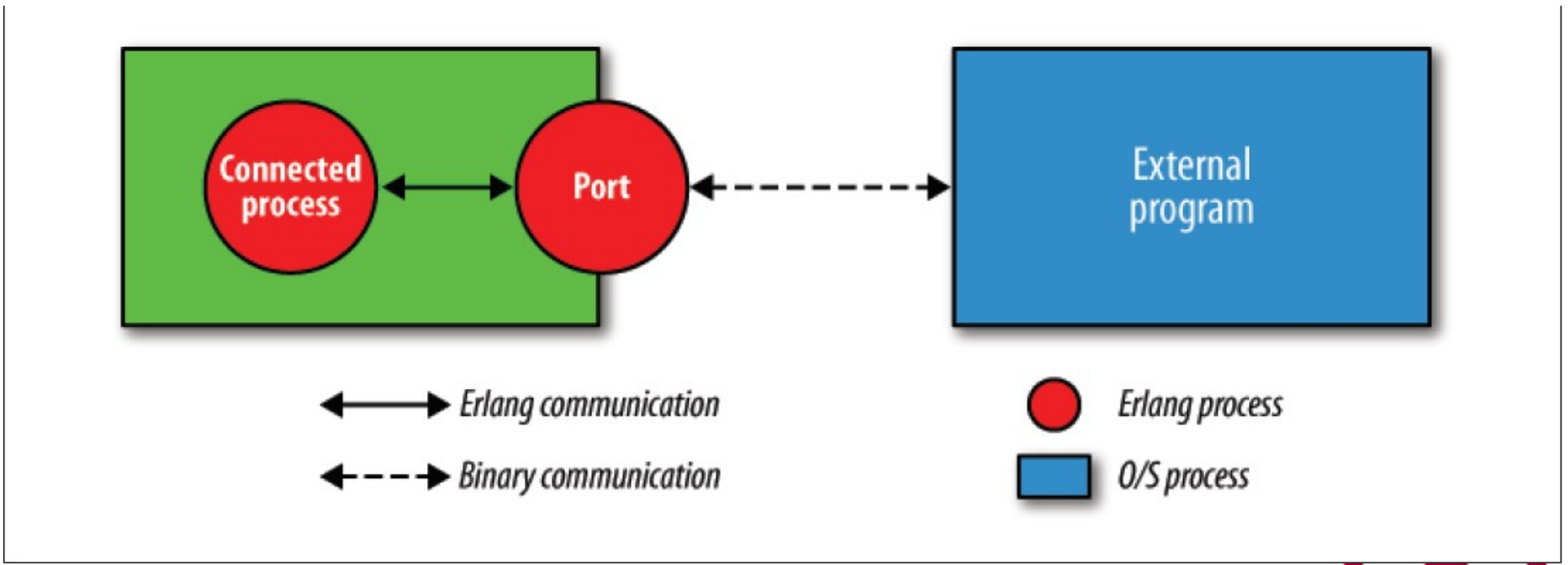
Concurrent Distributed FP Language

IO

- 对外的通讯通道
- POLL事件派遣机制
- 公平调度
- 透过系统管道连接异构系统

IO续

- Port负责对外的IO通信
- Port都有个宿主进程，用于协调通信



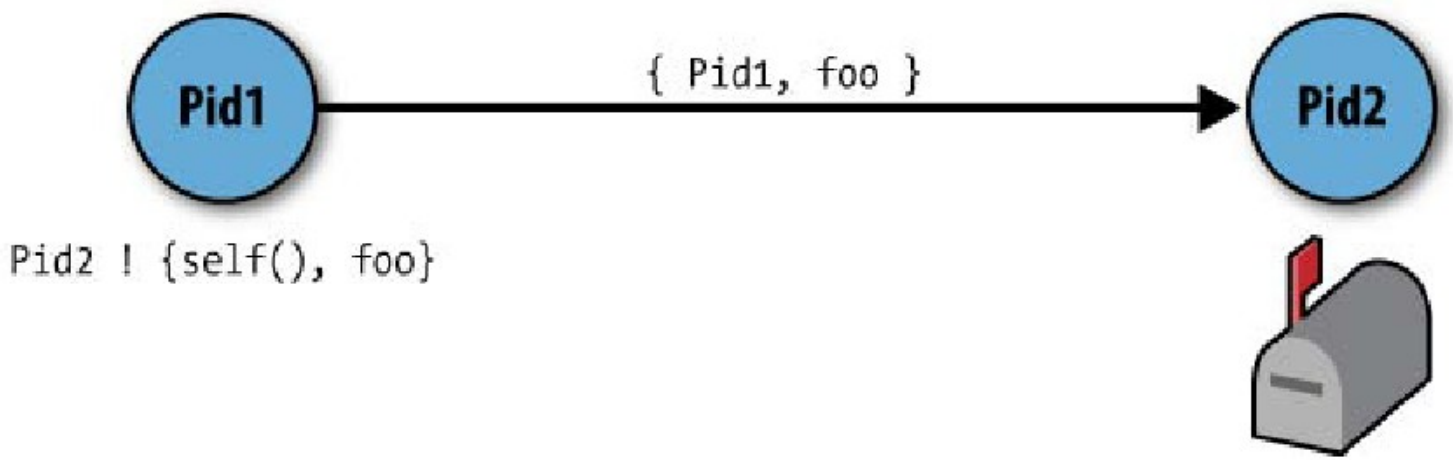
ERLANG
Concurrent Distributed FP Language

消息传递

- 消息传递是唯一的进程间通信方式
- 消息高效传递
- 消息透明传递
- 消息的跟踪
- 高效的消息编解码

消息续

- 进程级别的消息队列
- 选择性接收
- 超时



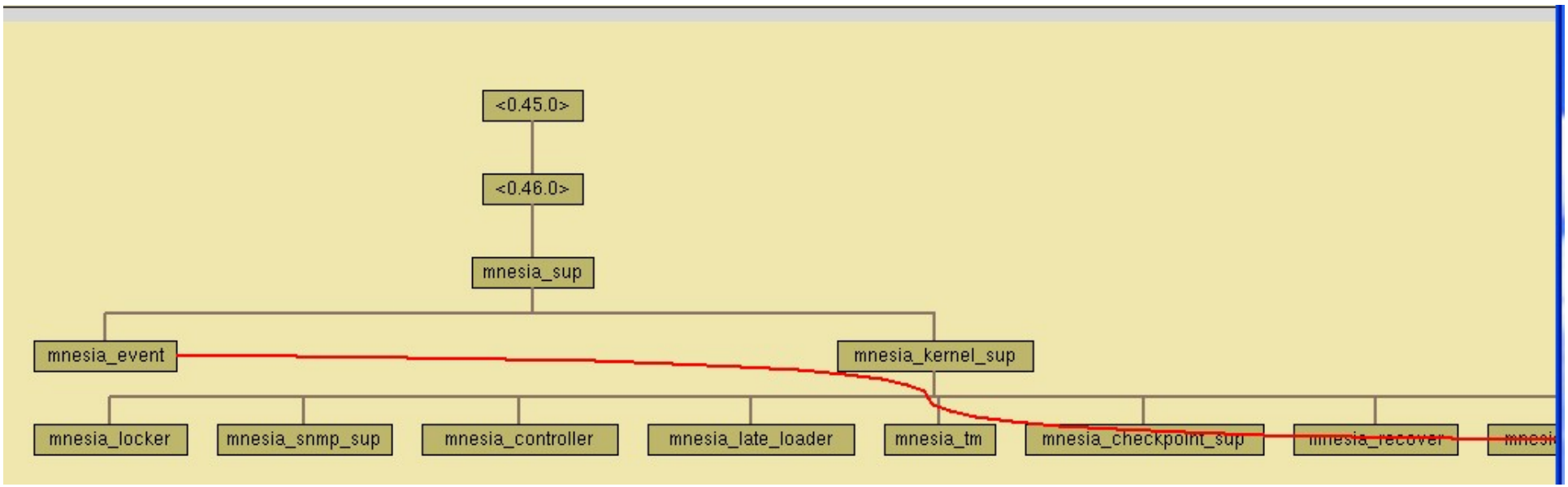
ERLANG
Concurrent Distributed FP Language

定时器

- 支持time jump detection and correction
- 定时器wheel实现，支持百万级别的定时器
- 多样的使用支持
  - 语法
  - 驱动
  - BIF

业务抽象

- 面向并发编程
- 轻量进程对应于现实世界的Actor
- Actor互动透过消息
- 模块提供基本的业务抽象



典型的一个应用有不同的进程, 充当不同的角色

ERLANG

Concurrent Distributed FP Language

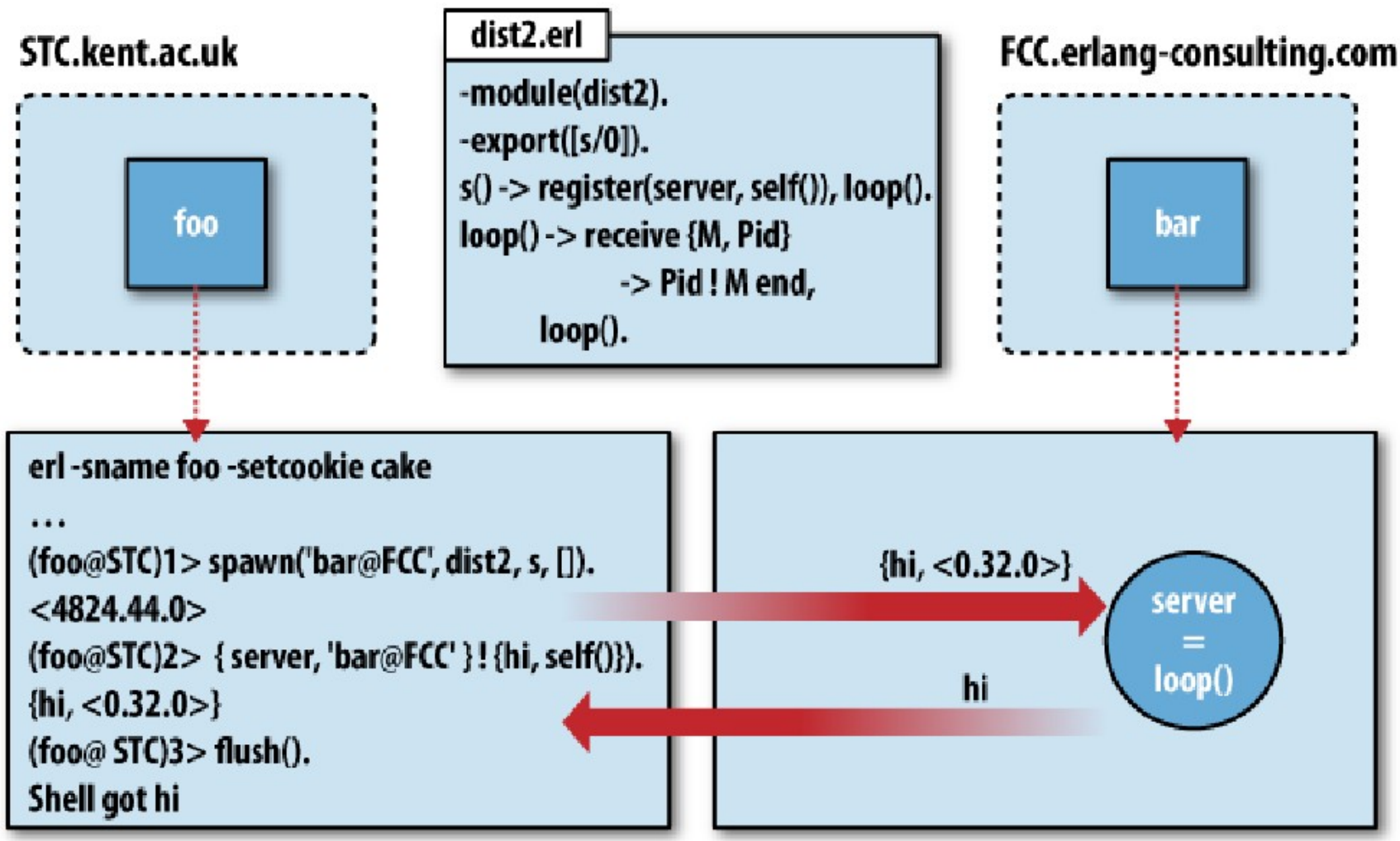
那么相比，Erlang VM的特点是

- 高效的数据结构
  - Atom, Binary, List, Tuple 4种基础数据
- GC Mark and sweep, 隔代, 进程级别
  - 软实时
- 资源自动回收
- 异步线程机制
  - 驱动可用
- 高效的内存分配
  - CPU独立内存池
  - 多种类型和类别的高速内存池，可微调
- 高效的锁机制
  - 高效的锁和检查机制

Erlang集群的设施

- Net_kernel, EPMD核心部件
- 可替换的传输介质
  - Inet_tcp_dist
  - Inet_ssl_dist
- group_leader的设计和用途
  - 截获输出
- dist trap 透明的进行握手动作
  - Connect and handshake
- 名称登记和维护
  - Local/global
- 维护网络全联通
  - Net tick
  - Nodeup nodedown

图例分布消息通信

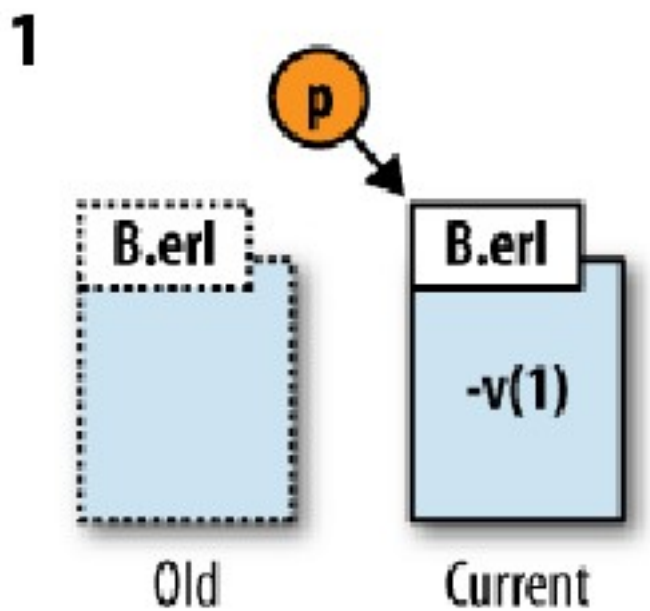


Erlang自省机制

- 巨强大的Trace机制和工具支持
- 巨完善的系统信息，可以不夸张的说可以获取任何信息
- 多样的监控
  - OS层面的cpu，memory，disk
  - VM层面的CPU，memory，GC活动等
- 规范的信息获取，每个模块都有个info函数
- CrashDump
- 完善的获取手段
  - Eshell
  - RPC
  - SSH
  - TOP/MAN等工具

代码的热升级

- 支持代码从**压缩包**和网络读取
- 代码热升级
  - Beam文件
  - 驱动层面(动态库)



ERLANG
Concurrent Distributed FP Language

内置强大的数据库

- ETS内存数据库
  - 支持Hash和Tree模式
  - 内置的虚拟机，执行高效的Match操作
  - 完善的接口
  - 不参与GC
- DETS持久数据库
  - 缺点只支持2G文件
- Mnesia集群数据库
  - 无中心节点的模式
  - 读写都在内存

稳定和移植性

- 稳定性
  - 号称6个9
  - Link/Monitor机制
  - 异常机制
  - 监督树
- 目前平台移植性:
  - Solaris
  - Linux
  - Freebsd
  - Mac
  - Windows

谢谢大家

联系我: mryufeng@gmail.com