

B4: Google's Software-Defined WAN

Paper Reading

Log Creative

2021 年 10 月 17 日

论文

Chi-Yao Hong et al. “B4 and after: Managing Hierarchy, Partitioning, and Asymmetry for Availability and Scale in Google’s Software-Defined WAN”. In: *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*. SIGCOMM ’18. Budapest, Hungary: Association for Computing Machinery, 2018, pp. 74–87. ISBN: 9781450355674. DOI: 10.1145/3230543.3230545. URL: <https://doi.org/10.1145/3230543.3230545>



B4

Google 私有广域网后端

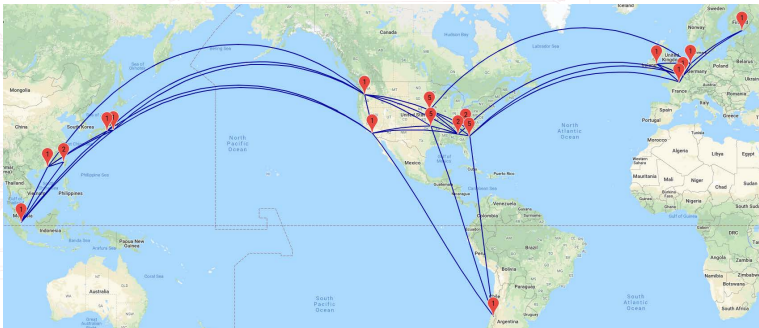
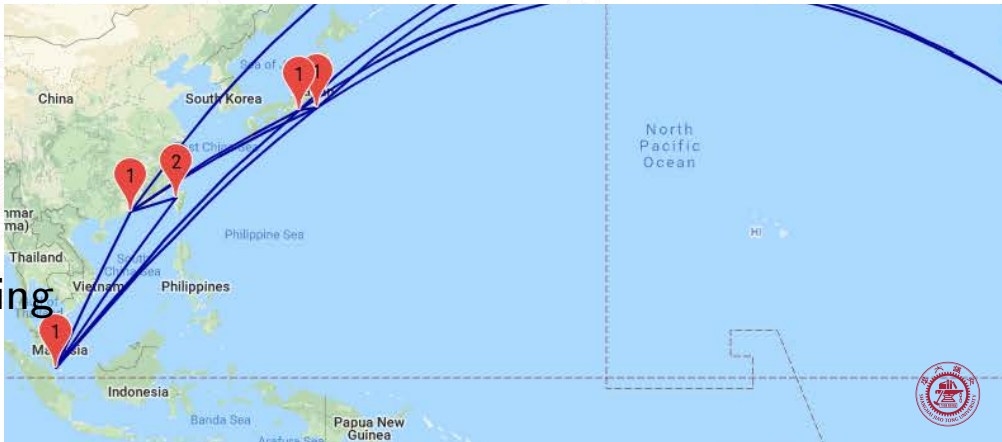


图: B4 全球网络



B4

Google 私有广域网后端

TE
Traffic
Engineering

SLO

Service Level Objectives 服务级别协议

表示 30 天滑动窗口内的网络连接可用性和带宽可用性。

服务级别	应用举例	SLO需求
SC4	搜索广告、DNS、WWW	99.99%
SC3	照片服务后端、邮件	99.95%
SC2	广告数据库拷贝	99.90%
SC1	搜索索引拷贝	99%
SC0	批量传输	

表: SLO



SLO

Service Level Objectives 服务级别协议

表示 30 天滑动窗口内的网络连接可用性和带宽可用性。

服务级别	应用举例	SLO需求
SC4	搜索广告、DNS、WWW	99.99%
SC3	照片服务后端、邮件	99.95%
SC2	广告数据库拷贝	99.90%
SC1	搜索索引拷贝	99%
SC0	批量传输	

表: SLO



SLO

Service Level Objectives 服务级别协议

表示 30 天滑动窗口内的网络连接可用性和带宽可用性。

服务级别	应用举例	SLO需求
SC4	搜索广告、DNS、WWW	99.99%
SC3	照片服务后端、邮件	99.95%
SC2	广告数据库拷贝	99.90%
SC1	搜索索引拷贝	99%
SC0	批量传输	

表: SLO



扁平结构

不利于扩展和可用性

之前的 B4 若想增加容量，需要在地理限界内增加站点。但这会带来：

- ① 增加了中央流量控制优化算法的运行时间。
- ② 对交换机有限的流表空间增加压力。
- ③ 使得容量管理变得复杂并给应用开发者造成麻烦。



扁平结构

不利于扩展和可用性

之前的 B4 若想增加容量，需要在地理限界内增加站点。但这会带来：

- ① 增加了中央流量控制优化算法的运行时间。
- ② 对交换机有限的流表空间增加压力。
- ③ 使得容量管理变得复杂并给应用开发者造成麻烦。

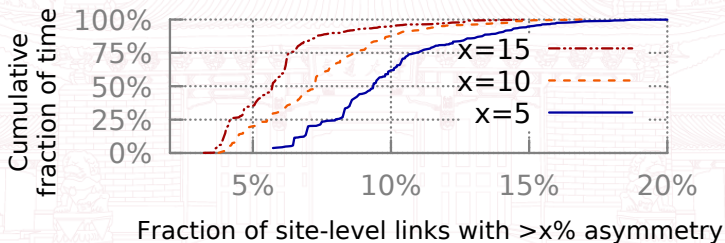
为了解决这个问题，引入 *supernode* 和两层架构。



层级架构

容量不对等问题

B4 中 6–20% 的地理级连接仍然会在 $\geq 5\%$ 的时间内有容量不对等情形。



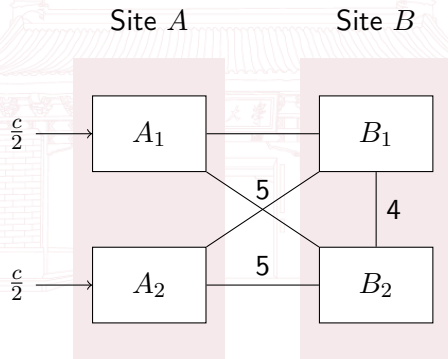
$$\frac{\text{avg}_{\forall i} C_i - \min_{\forall i} C_i}{\text{avg}_{\forall i} C_i}$$

图：地理级流量不对等



不对等的后果

大幅减少系统效率

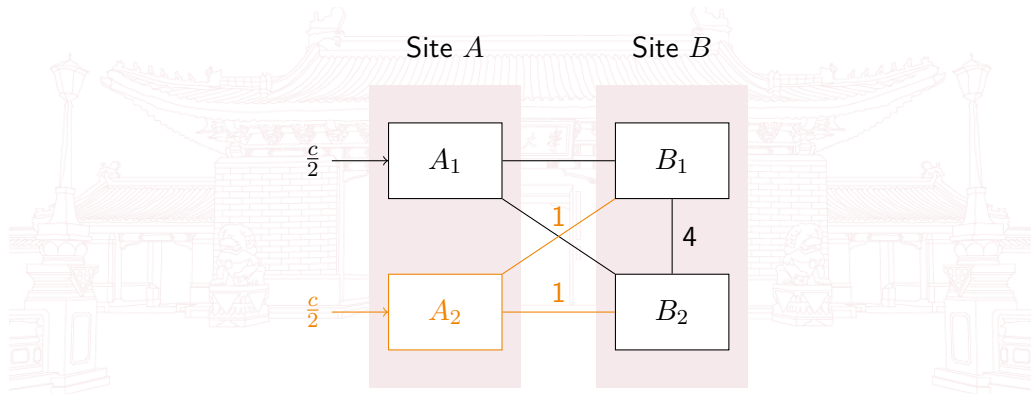


图：对等



不对等的后果

大幅减少系统效率

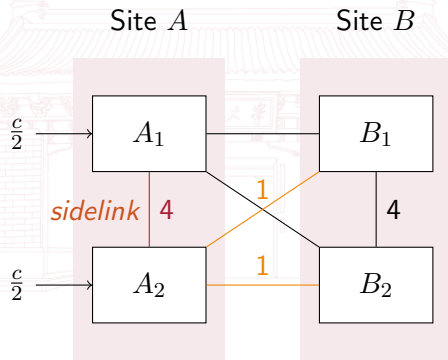


图：不对等示例 $c = 4$



不对等的后果

大幅减少系统效率



图：不对等示例 $c = 12$



使用 *sidelink* 可以提高不对等时的带宽利用率。但是仍然需要考虑相关的协议问题，比如有些数据不可分割、MAC 地址不可变化，以及死循环问题，转换隧道可能是原子操作，以任意顺序应用 TE 更新会导致这种死循环率上升，



高效交换规则管理

Merchant 交换机只支持有限的匹配和哈希规则。



Saturn

第一代 B4 网络结构

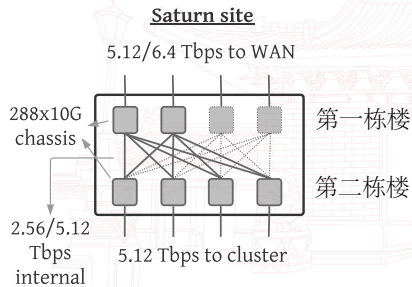


图: Saturn 站点

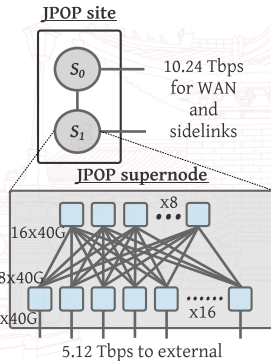
名称	Saturn
部署年	2010
类型	数据中心
交换机芯片	24x10G
每站点机箱数	6 / 8
站点容量 (Tbps)	5.12 EX 2.56 INTER
每站点交换机箱数	4
控制域数量	1

表: Saturn 站点



Jumpgate: JPOP

仅传输站点



名称	JPOP
部署年	2013
类型	POP
交换机芯片	16x40G
每站点机箱数	20
超级节点交换机数	24
站点容量 (Tbps)	10.24
每站点交换机箱数	4
控制域数量	2

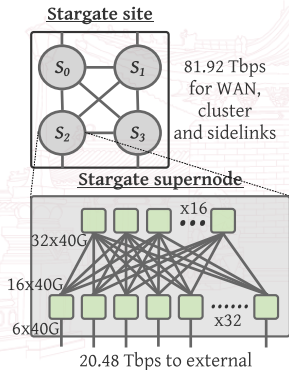
图: JPOP 站点

表: JPOP 站点



Jumpgate: Stargate

数据中心级



名称	Stargate
部署年	2014
类型	数据中心
交换机芯片	32x40G
每站点机箱数	192
超级节点交换机数	48
站点容量 (Tbps)	81.92
每站点交换机箱数	8
控制域数量	4

表: Stargate 站点

图: Stargate 站点



Jumpgate: Stargate

数据吞吐量大带来的好处

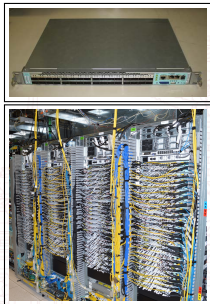


图: 交换机与交换机箱

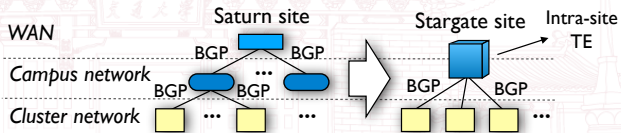


图: 减少 BGP 复杂度





谢谢