

NGINX

L4/L7 Load Balance

臺大計資中心網路組
游子興

davisyou@ntu.edu.tw
02-33665008

大綱

- * Load-Balancing
- * L4 Load Balance
 - * TCP
 - * UDP
- * L7 Load Balance

Load Balance 參考文件

- * NGINX

- * TCP and UDP Load Balancing

- * <https://docs.nginx.com/nginx/admin-guide/load-balancer/tcp-udp-load-balancer/>

- * HTTP Load Balancing

- * <https://docs.nginx.com/nginx/admin-guide/load-balancer/http-load-balancer/>

- * <https://www.nginx.com/products/nginx/load-balancing/>

- * What are the best load balancing methods and algorithms?

- * <https://www.loadbalancer.org/blog/load-balancing-methods/>

- * Others

- * <https://www.loadbalancer.org/blog/how-to-achieve-client-source-ip-transparency-with-loadbalancer-org/>

- * <https://www.loadbalancer.org/blog/nginx-and-x-forwarded-for-header/>

Load-Balance Methods

- * Round Robin(default)
- * Least Connections
 - * least_conn;
- * Least Time (NGINX Plus only)
 - * Based on a calculation that combines response time and number of active connections.
 - * least_time first_byte;
 - * least_time header;
- * Hash
 - * Based on a specified key.
 - * hash \$remote_addr;
- * IP Hash(HTTP only)
 - * Based on the first three octets of IPv4 or entire IPv6 address.
 - * ip_hash;
- * Random
 - * random two least_time=last_byte;

For 每台 Server 個別設定 server address [parameters];

- * weight=1(default)
 - * 越大優先權越高.
- * max_conns=0(default) no limit
 - * limits the maximum number of simultaneous active connections
- * max_fails=1(default)
 - * the number of fail connections in the fail_timeout duration to consider the server unavailable.
- * fail_timeout=10(default) seconds.
- * backup
 - * work when the primary servers are unavailable.
- * down
 - * marks unavailable.

-
- * proxy_connect_timeout 60s(default);
 - * a timeout for establishing a connection with a proxied server

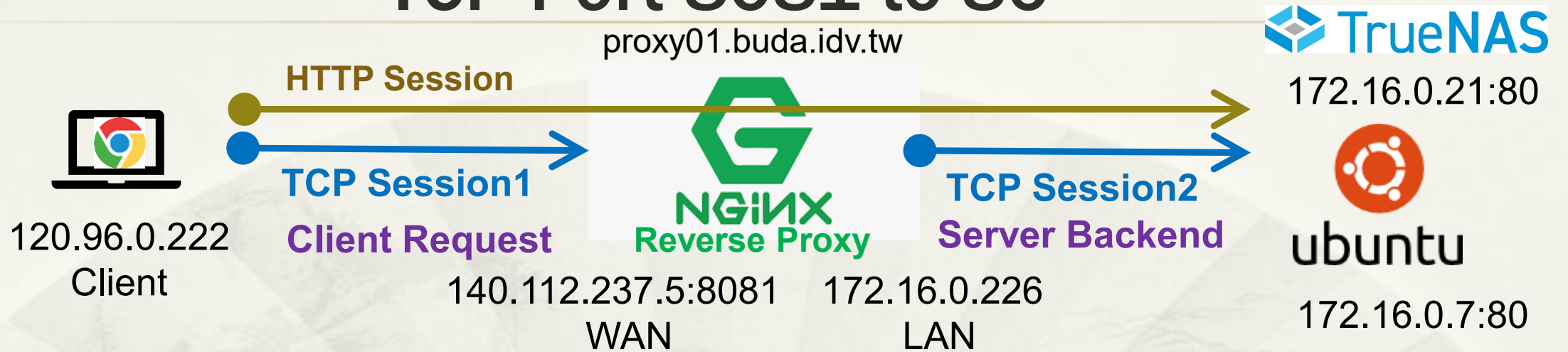


L4 Load Balance

Round Robin

L4 Load Balance

TCP Port 8081 to 80



使用者連線:

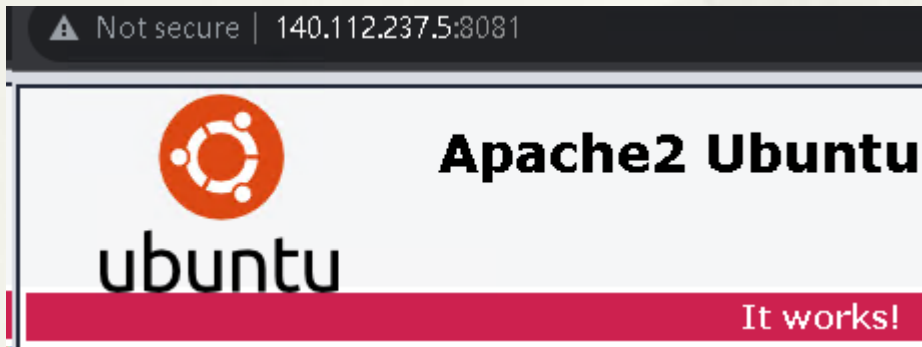
Test1: <http://140.112.237.5:8081/>

Test2: <http://proxy01.buda.idv.tw:8081/>

```
* /etc/nginx/nginx.conf
stream {
    server {
        listen 8081;
        #proxy_connect_timeout 10s;
        proxy_pass backend_tcp_80;
    }
    upstream backend_tcp_80 {
        server 172.16.0.4:80;
        server 172.16.0.7:80;
    }
}
```

L4 Load Balance (TCP)

測試 1



Browser 有 cache 功能, 測試不準確.
建議用 `curl http://140.112.237.5:8081`

L4 Load Balance (TCP)

Server Down 測試

中斷網卡連線



斷線

15:13:42.345332 IP ubuntu2204.40062 > 172.16.0.7.http: Flags [S], seq 1124550674, win 64240, options [mss 1460,sackOK,TS val 3821693597 ecr 0,nop,wscale 7], length 0
15:13:43.385887 IP ubuntu2204.58950 > 172.16.0.1.http: Flags [S], seq 3982080893, win 64240, options [mss 1460,sackOK,TS val 1945032026 ecr 0,nop,wscale 7], length 0
15:13:44.433519 IP ubuntu2204.40076 > 172.16.0.7.http: Flags [S], seq 2422243388, win 64240, options [mss 1460,sackOK,TS val 3821695685 ecr 0,nop,wscale 7], length 0
15:13:45.453797 IP ubuntu2204.40076 > 172.16.0.7.http: Flags [S], seq 2422243388, win 64240, options [mss 1460,sackOK,TS val 3821696706 ecr 0,nop,wscale 7], length 0
15:13:47.469774 IP ubuntu2204.40076 > 172.16.0.7.http: Flags [S], seq 2422243388, win 64240, options [mss 1460,sackOK,TS val 3821698722 ecr 0,nop,wscale 7], length 0
15:13:51.693767 IP ubuntu2204.40076 > 172.16.0.7.http: Flags [S], seq 2422243388, win 64240, options [mss 1460,sackOK,TS val 3821702946 ecr 0,nop,wscale 7], length 0
15:13:59.885789 IP ubuntu2204.40076 > 172.16.0.7.http: Flags [S], seq 2422243388, win 64240, options [mss 1460,sackOK,TS val 3821711138 ecr 0,nop,wscale 7], length 0
15:14:16.017772 IP ubuntu2204.40076 > 172.16.0.7.http: Flags [S], seq 2422243388, win 64240, options [mss 1460,sackOK,TS val 3821727270 ecr 0,nop,wscale 7], length 0
15:14:44.454000 IP ubuntu2204.54686 > 172.16.0.1.http: Flags [S], seq 1330842856, win 64240, options [mss 1460,sackOK,TS val 1945093094 ecr 0,nop,wscale 7], length 0
15:14:45.460397 IP ubuntu2204.54700 > 172.16.0.1.http: Flags [S], seq 3929634236, win 64240, options [mss 1460,sackOK,TS val 1945094100 ecr 0,nop,wscale 7], length 0

切換

預設需 60 秒

proxy_connect_timeout 60s(default);

L7 Load Balance

留意限制使用特定 Host Name 連線之網站
(WebServer 綁定 Virtual Host Name)

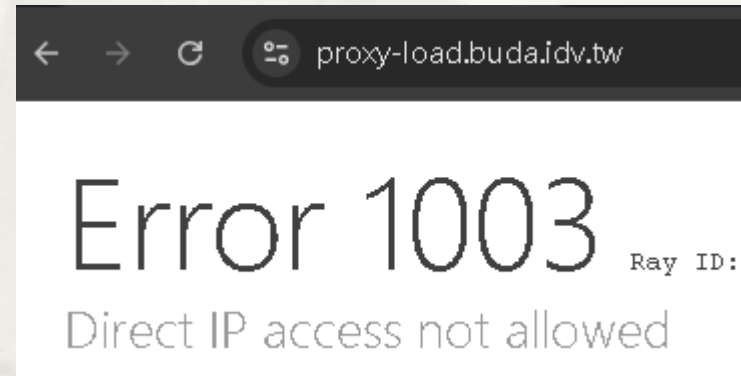
L7 Load Balance (HTTP to HTTP)

http://httpforever.com

```
http {  
    server {  
        listen 80;  
        server_name proxy-load.buda.idv.tw;  
        location / {  
            proxy_pass http://backend_http;  
        }  
    }  
    upstream backend_http {  
        server httpforever.com;  
    }  
}
```

* 使用者連線:

- * http://proxy-load.buda.idv.tw



* 連線失敗

L7_HTTP_LoadBalance.pcap

失敗原因

* tcp.stream eq 0

* tcp.stream eq 2

```
Wireshark · Follow TCP Stream (tcp.stream eq 0) · L7_HTTP_LoadBalance.pcap

GET / HTTP/1.1
Host: demo1.buda.idv.tw
Connection: keep-alive
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/116.0.0.0 Safari/537.36
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.7
Accept-Encoding: gzip, deflate
Accept-Language: zh-TW,zh;q=0.9

HTTP/1.1 400 Bad Request
Server: nginx/1.24.0
Date: Sun, 27 Aug 2023 13:40:27 GMT
Content-Type: text/html; charset=iso-8859-1
Content-Length: 347
Connection: keep-alive
```

```
Wireshark · Follow TCP Stream (tcp.stream eq 1) · L7_HTTP_LoadBalance.pcap

GET / HTTP/1.0
Host: backend_http
Connection: close
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/116.0.0.0 Safari/537.36
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.7
Accept-Encoding: gzip, deflate
Accept-Language: zh-TW,zh;q=0.9

HTTP/1.1 400 Bad Request
Date: Sun, 27 Aug 2023 13:40:27 GMT
Server: Apache/2.4.6 (CentOS) mod_auth_gssapi/1.5.1 mod_nss/1.0.14 NSS/3.28.4 mod_wsgi/3.4 Python/2.7.5 PHP/7.0.18
Content-Length: 347
Connection: close
Content-Type: text/html; charset=iso-8859-1
```

使用錯誤 Virtual Host Name 連線

L7_HTTP_LoadBalance.pcap

失敗原因

No.	Time	tcp.stream	TTL	Source	Src Port	Destination	Dest Port	Protocol	Length	Info
1	0.000000	0	126	140.112.3.82	25696	140.112.237.5	80	HTTP	486	GET / HTTP/1.1
2	0.000442	1	64	140.112.237.5	53906	140.112.51.148	80	HTTP	488	GET / HTTP/1.0
3	0.001082	1	59	140.112.51.148	80	140.112.237.5	53906	HTTP	679	HTTP/1.1 400 Bad Request (text/html)
4	0.001176	0	64	140.112.237.5	80	140.112.3.82	25696	HTTP	578	HTTP/1.1 400 Bad Request (text/html)
5	0.036314	0	126	140.112.3.82	25696	140.112.237.5	80	HTTP	432	GET /favicon.ico HTTP/1.1
6	0.036735	2	64	140.112.237.5	53920	140.112.51.148	80	HTTP	434	GET /favicon.ico HTTP/1.0
7	0.037132	2	59	140.112.51.148	80	140.112.237.5	53920	HTTP	679	HTTP/1.1 400 Bad Request (text/html)
8	0.037231	0	64	140.112.237.5	80	140.112.3.82	25696	HTTP	578	HTTP/1.1 400 Bad Request (text/html)

```
> Frame 1: 486 bytes on wire (3888 bits)
> Ethernet II, Src: Cisco_f8:21:08 (8c:94:08:f8:21:08), Dst: 02:00:0c:00:00:00
> Internet Protocol Version 4, Src: 140.112.3.82, Dst: 140.112.237.5
> Transmission Control Protocol, Src Port: 25696, Dst Port: 80
> Hypertext Transfer Protocol
  > GET / HTTP/1.1\r\n
  Host: demo1.buda.idv.tw\r\n
  Connection: keep-alive\r\n
```

```
> Frame 2: 488 bytes on wire (3904 bits)
> Ethernet II, Src: VMware_a6:67:9c (08:00:27:9c:a6:67), Dst: 02:00:0c:00:00:00
> Internet Protocol Version 4, Src: 140.112.237.5, Dst: 140.112.51.148
> Transmission Control Protocol, Src Port: 53906, Dst Port: 80
> Hypertext Transfer Protocol
  > GET / HTTP/1.0\r\n
  Host: backend_http\r\n
  Connection: close\r\n
```

L7 Load Balance (HTTP to HTTP)

httpforever.com 解決方法 1

```
http {  
    server {  
        listen 80;  
        server_name proxy-load.buda.idv.tw;  
        proxy_set_header Host "httpforever.com";  
        location / {  
            proxy_pass http://backend_http;  
        }  
    }  
    upstream backend_http {  
        server httpforever.com;  
    }  
}
```

* 解決

- * 直接設定 Header Host

L7 Load Balance (HTTP to HTTP)

httpforever.com 解決方法 2

```
http {  
    server {  
        listen 80;  
        server_name proxy-load.buda.idv.tw;  
        location / {  
            proxy_pass http://httpforever.com;  
        }  
    }  
    upstream httpforever.com {  
        server httpforever.com;  
    }  
}
```

- * 解決
 - * Load Balance upstream 直接用 Virtual Host Name
- * 考慮順序
 - * NGINX 內部 httpforever.com 設定
 - * 高於對應外部網站 Domain Name

L7_HTTP_LoadBalance2.pcap

tcp.stream eq 0

Wireshark · Follow TCP Stream (tcp.stream eq 0) · L7_HTTP_LoadBalance2.pcap

```
GET / HTTP/1.1
Host: demo1.buda.idv.tw
Connection: keep-alive
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML
L, like Gecko) Chrome/116.0.0.0 Safari/537.36
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image
/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.7
Accept-Encoding: gzip, deflate
Accept-Language: zh-TW,zh;q=0.9
```

```
HTTP/1.1 200 OK
Server: nginx/1.24.0
Date: Sun, 27 Aug 2023 13:59:06 GMT
Content-Type: text/html; charset=utf-8
Transfer-Encoding: chunked
Connection: keep-alive
X-Content-Type-Options: nosniff
X-Powered-By: PHP/7.0.18
X-Drupal-Cache: HIT
Etag: "1693137910-1"
Content-Language: zh-hant
X-Frame-Options: SAMEORIGIN
X-Content-Type-Options: nosniff
X-Generator: Drupal 7 (http://drupal.org)
Cache-Control: public, max-age=1800
```

tcp.stream eq 1

Wireshark · Follow TCP Stream (tcp.stream eq 1) · L7_HTTP_LoadBalance2.pcap

```
GET / HTTP/1.0
Host: www.math.ntu.edu.tw 正確 Virtual Host Name
Connection: close
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML
L, like Gecko) Chrome/116.0.0.0 Safari/537.36
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image
/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.7
Accept-Encoding: gzip, deflate
Accept-Language: zh-TW,zh;q=0.9
```

```
HTTP/1.1 200 OK
Date: Sun, 27 Aug 2023 13:59:06 GMT
Server: Apache/2.4.6 (CentOS) mod_auth_gssapi/1.5.1 mod_nss/1.0.14 NSS/3.28.4
mod_wsgi/3.4 Python/2.7.5 PHP/7.0.18
X-Content-Type-Options: nosniff
X-Powered-By: PHP/7.0.18
X-Drupal-Cache: HIT
Etag: "1693137910-1"
Content-Language: zh-hant
X-Frame-Options: SAMEORIGIN
X-Content-Type-Options: nosniff
X-Generator: Drupal 7 (http://drupal.org)
Cache-Control: public, max-age=1800
Last-Modified: Sun, 27 Aug 2023 12:05:10 GMT
Expires: Sun, 19 Nov 1978 05:00:00 GMT
```

L7_HTTP_LoadBalance2.pcap

No.	Time	tcp.stream	TTL	Source	Src Port	Destination	Dest Port	Protocol	Length	Info
1	0.000000	0	126	140.112.3.82	61934	140.112.237.5	80	TCP	66	61934 → 80 [SYN] Seq=0 Win=64240 Len=0 MSS=9
2	0.000022	0	64	140.112.237.5	80	140.112.3.82	61934	TCP	66	80 → 61934 [SYN, ACK] Seq=0 Ack=1 Win=64240
3	0.000217	0	126	140.112.3.82	61934	140.112.237.5	80	TCP	60	61934 → 80 [ACK] Seq=1 Ack=1 Win=262912 Len=
4	0.000346	0	126	140.112.3.82	61934	140.112.237.5	80	HTTP	486	GET / HTTP/1.1
5	0.000354	0	64	140.112.237.5	80	140.112.3.82	61934	TCP	54	80 → 61934 [ACK] Seq=1 Ack=433 Win=64128 Len=
6	0.000507	1	64	140.112.237.5	51456	140.112.51.148	80	TCP	74	51456 → 80 [SYN] Seq=0 Win=64240 Len=0 MSS=1
7	0.000786	1	59	140.112.51.148	80	140.112.237.5	51456	TCP	74	80 → 51456 [SYN, ACK] Seq=0 Ack=1 Win=28960
8	0.000794	1	64	140.112.237.5	51456	140.112.51.148	80	TCP	66	51456 → 80 [ACK] Seq=1 Ack=1 Win=64256 Len=6
9	0.000822	1	64	140.112.237.5	51456	140.112.51.148	80	HTTP	495	GET / HTTP/1.0
10	0.001038	1	59	140.112.51.148	80	140.112.237.5	51456	TCP	66	80 → 51456 [ACK] Seq=1 Ack=430 Win=30080 Len=
11	0.006934	1	59	140.112.51.148	80	140.112.237.5	51456	TCP	1514	80 → 51456 [ACK] Seq=1 Ack=430 Win=30080 Len=

```
> Frame 4: 486 bytes on wire (388
> Ethernet II, Src: Cisco_f8:21:c
> Internet Protocol Version 4, Sr
> Transmission Control Protocol,
✓ Hypertext Transfer Protocol
  > GET / HTTP/1.1\r\n
    Host: demo1.buda.idv.tw\r\n
    Connection: keep-alive\r\n
```

```
> Frame 9: 495 bytes on wire (3960
> Ethernet II, Src: VMware_a6:67:c
> Internet Protocol Version 4, Src
> Transmission Control Protocol, S
✓ Hypertext Transfer Protocol
  > GET / HTTP/1.0\r\n
    Host: www.math.ntu.edu.tw\r\n
    Connection: close\r\n
```

Sticky Sessions (Session Cookie Persistence)

Client 固定分配 to Specific Server

Sticky Sessions

- * NGINX Open Source
 - * hash or ip_hash
 - * 缺點
 - * NAT Pool:
 - * Dynamic IPs: Mobile clients changing IP addresses mid-session lose their session context.
- * NGINX Plus supports 3 persistence methods.
 - * Sticky cookie (Cookie Insert)
 - * Sticky learn (Cookie Learn)
 - * Sticky route

Sticky cookie

- * adds a session cookie to the first response from the upstream group and identifies the server that sent the response.
- * The client's next request contains the cookie value and route the request to the upstream server that responded to the first request.
- * Config

```
upstream backend {  
    sticky cookie srv_id expires=1h domain=.example.com path=/  
    server backend1.example.com;  
    server backend2.example.com;  
}
```

Sticky learn

- * 1. Finds session identifiers by inspecting requests and responses.
- * 2. “learns” which upstream server corresponds to which session identifier.
- * Generally, these identifiers are passed in a HTTP cookie. If a request contains a session identifier already “learned”, NGINX Plus forwards the request to the corresponding server:

- * Config

```
upstream backend {  
    sticky learn  
        create=$upstream_cookie_examplecookie  
        lookup=$cookie_examplecookie  
        zone=client_sessions:1m  
        timeout=1h;  
  
    server backend1.example.com;  
    server backend2.example.com;  
}
```

Sticky route

- * assigns a “route” to the client when it receives the first request.
- * All subsequent requests are compared to the route parameter of the server directive to identify the server to which the request is proxied.
- * The route information is taken from either a cookie or the request URI.
- * Config

```
upstream backend {  
    sticky route $route_cookie $route_uri;  
    server backend1.example.com route=a;  
    server backend2.example.com route=b;  
}
```

L7 Load Balance (HTTPs to HTTPs)

```
http {  
    server {  
        listen 443 ssl;  
        server_name proxy-load.buda.idv.tw;  
        location / {  
            proxy_pass http://backend_https;  
        }  
    }  
    upstream backend_http {  
        #ip_hash;  
        server 172.16.0.26:443;  
        server 172.16.0.25:443;  
        server 172.16.0.21:443;  
    }  
}
```

- * 連線測試
- * https://172.16.0.21/php/login/login_ok.php
- * https://172.16.0.25/php/login/login_ok.php
- * https://172.16.0.26/php/login/login_ok.php
- * https://proxy-load.buda.idv.tw/php/login/login_ok.php
- * User: user01
- * Passwd: 54321



簡報完畢
謝謝