<http://wangcong.org/blog/archives/1683>

[**各種tunnel**](http://wangcong.org/blog/archives/1683)

網絡中有各種各樣的[tunnel](http://en.wikipedia.org/wiki/Tunneling_protocol)。tunnel是把下一層(如IPv4層)的包封裝到上一層(如SSH，HTTP)或同一層(如IPv6層)的協議中進行傳輸，實現網絡之間的穿透。發送端和接收端須各有一個解析這種包的程序或者內核模塊才能通信。

最簡單的例子，SSH Tunnel。在本地要做端口映射，在中轉的ssh服務器上要做port forward。把HTTP做到SSH上，算是[TCP over TCP](http://sites.inka.de/bigred/devel/tcp-tcp.html)。這是應用層的tunnel。內核中的tunnel有四個：ipip、gre、sit、isatap。通過ip tunnel 命令看到：

% ip tunnel help

Usage: ip tunnel { add | change | del | show | prl | 6rd } [ NAME ]

[ mode { ipip | gre | sit | isatap } ] [ remote ADDR ] [ local ADDR ]

[ [i|o]seq ] [ [i|o]key KEY ] [ [i|o]csum ]

[ prl-default ADDR ] [ prl-nodefault ADDR ] [ prl-delete ADDR ]

[ 6rd-prefix ADDR ] [ 6rd-relay\_prefix ADDR ] [ 6rd-reset ]

[ ttl TTL ] [ tos TOS ] [ [no]pmtudisc ] [ dev PHYS\_DEV ]

...

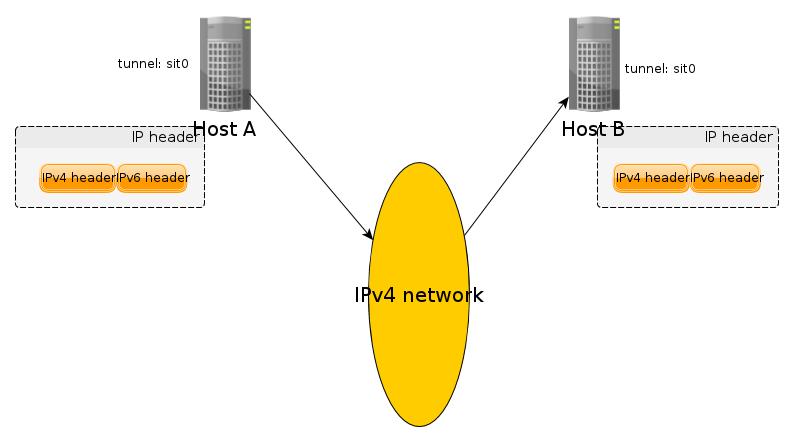
ipip是把IP層封裝到IP層的一個tunnel，它的作用相當於一個基於IP層的網橋！普通的網橋是基於mac層的，不需IP，而ipip則通過兩端的路由做一個tunnel，把兩個本來不通的網絡通過點對點連接起。

gre和它類似，但它功能還要更強大一些，還支持廣播，可以取代ipip。

<http://wangcong.org/blog/archives/2149>

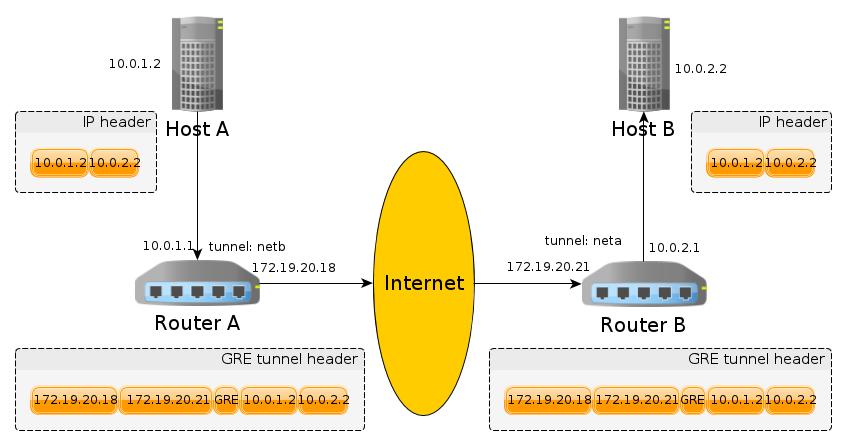
## [深入理解GRE tunnel](http://wangcong.org/blog/archives/2149)

多數tunnel容易理解，因為多是一對一的，換句話說，是直接從一端到另一端。比如IPv6 over IPv4的tunnel，也就是SIT，它的原理如下圖所示：

[](http://wangcong.org/blog/wp-content/uploads/2012/11/sit_tunnel1.jpg)

除了端點的host A和host B之外，中間經過的任何設備都是看不到裡面的IPv6的頭，對於它們來說，經過sit發出的包和其它的IPv4的包沒有任何區別。

[GRE tunnel](http://en.wikipedia.org/wiki/Generic_Routing_Encapsulation)的根本原理和sit，ipip不一樣。​​外層的IP頭和內層的IP頭之間多了一個GRE頭之外，它最大的不同是**tunnel不是建在最終的host上，而是在中間的router上**！換句話說，對於端點host A和host B來說，該tunnel是透明的（對比上面的sit tunnel，所以它才能解決ipip tunnel解決不了的問題。所以經過GRE tunnel發送的包(從host A發送到host B)大體過程是這樣子的：

[](http://wangcong.org/blog/wp-content/uploads/2012/11/gre_tunnel1.jpg)

從host A發出的包是一個普通的IP包，目的地址不可直接到達。該GRE tunnel的一端是建立在router A上，另一段是建立在router B上，所以添加外部的IP 頭是在router A上完成的，而去掉外面的IP頭是在router B上完成的，兩個端點的host上幾乎什麼都不用做(除了配置路由，把發送到10.0.2.0 的包路由到router A)！

這麼設計的好處很容易看出來，ipip tunnel是端對端的，通信只能點對點，GRE tunnel卻可多播。

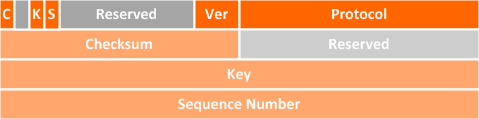
發送過程中，router A上配置了一條路由規則，凡是發往10.0.2.0網絡的包都要經過構造一個新的頭，並把對應的外部IP地址填進去，最後發送出去。

接收過程中，即router B進行的操作。GRE tunnel自己定義一個新的IP proto，也就是IPPROTO\_GRE。當router B收到從router A過來的包時，暫時還不知道這是GRE的包，它首先會當作普通的IP包處理。因為外部的IP頭的目的地址是該路由器的地址，所以它自己會接收這個包，把它交給上層，到了IP層後發現這個包不是TCP，UDP，而是GRE，這時內核會轉交給GRE模塊處理。通過外層IP頭找到對應的tunnel，然後剝去外層IP 頭，把這個“新的”包重新交給IP棧去處理，像接收到普通IP包一樣。到了這裡，“新的”包處理和其它普通的IP包已經沒什麼兩樣了：根據IP頭中目的地址轉發給相應的host。注意：“新的”包其實並不新，內核用的還是同一個copy，只是skbuff 里相應的指針變了。

IP header in GRE tunnel

[http://wangcong.org/blog/wp-content/uploads/2012/11/gre_encapsulation.png](http://wangcong.org/blog/wp-content/uploads/2012/11/gre_encapsulation.png)

GRE header

[](http://wangcong.org/blog/wp-content/uploads/2012/11/gre_header.png)