Lọc dữ liệu

- Sử dụng nhiều hệ đo.
- Sử dụng giá trung bình của một số lớn phép đo liên tục.
- Loại bỏ các số liệu không đáng tin (sai số lớn):
  - Dựa trên phân tích xu hướng.
  - Dựa trên sai số trung bình nhiều lần đo.
  - Dựa trên sai số so với dự báo.
- Sử dụng các cảm biến dựa trên các phương pháp đo khác nhau để kiểm định và chuẩn hóa định kỳ.

Lưu trữ dữ liệu

- Sử dụng các CSDL (SQL):
  - MySQL
  - MariaDB
  - PostgreSQL
  - mongoDB
- Sử dụng file - lưu trữ tạm.
  - Trên linux: /sys/class/thermal/thermal_zone0/temp

Lưu trữ dữ liệu

Các dữ liệu đo đạc của IoT từ các cảm biến đều có dạng bảng - bao gồm 2 cột thời gian đo và giá trị đo. Vì vậy các dữ liệu này được lưu trữ chủ yếu trên các CSDL dạng SQL (Structure Query Language). Đây là một ngôn ngữ được công bố bởi các RFC.

Có rất nhiều các chương trình (program) đang được sử dụng tuân thủ ít hoặc nhiều, thậm chí đầy đủ theo RFC.

Ví dụ SQL (MySQL / MariaDB)

- Tạo USER: CREATE USER 'quangnh'@'localhost' IDENTIFIED BY 'IoT_0123';
- Tạo CSDL: CREATE DATABASES hus_moodle;
- Giao quyền: GRANT ALL PRIVILEGES ON hus_moodle. * TO “quangnh''@'localhost';
- Liệt kê CSDL: SHOW DATABASES;
- Chọn CSDL: USE hus_moodle;
- Liệt kê bảng: SHOW TABLES;
- Lấy dữ liệu: SELECT usernname FROM user;

  SELECT * FROM user WHERE username=’quangnh’;

  UPDATE user SET password=”1234” WHERE username=’quangnh‘;

  DELETE FROM username WHERE username=’quangnh‘;

Phân tích dữ liệu

- Phân tích thống kê: Sử dụng phương pháp thống kê để phân tích dữ liệu. Phương pháp này không trực tiếp chỉ ra được các nguyên nhân tạo nên sự biến đổi của số liệu. Phương pháp này có thể cho chúng ta thấy sự tương quan giữa các số liệu hay chỉ ra cho chúng ta thấy một “giả định” nào đó có thể chấp nhận được hay không.
- Phân tích mô hình - sử dụng các nguyên lý động lực học: là phương pháp sử dụng các nguyên lý động lục học để phân tích số liệu. Dự báo thời tiết là một ví dụ điển hình của phương pháp này.
- Big data: là một lĩnh vực mới chủ yếu được sử dụng cho các số liệu xã hội học, như kinh tế chẳng hạn.

An toàn và bảo mật hệ thống IoT

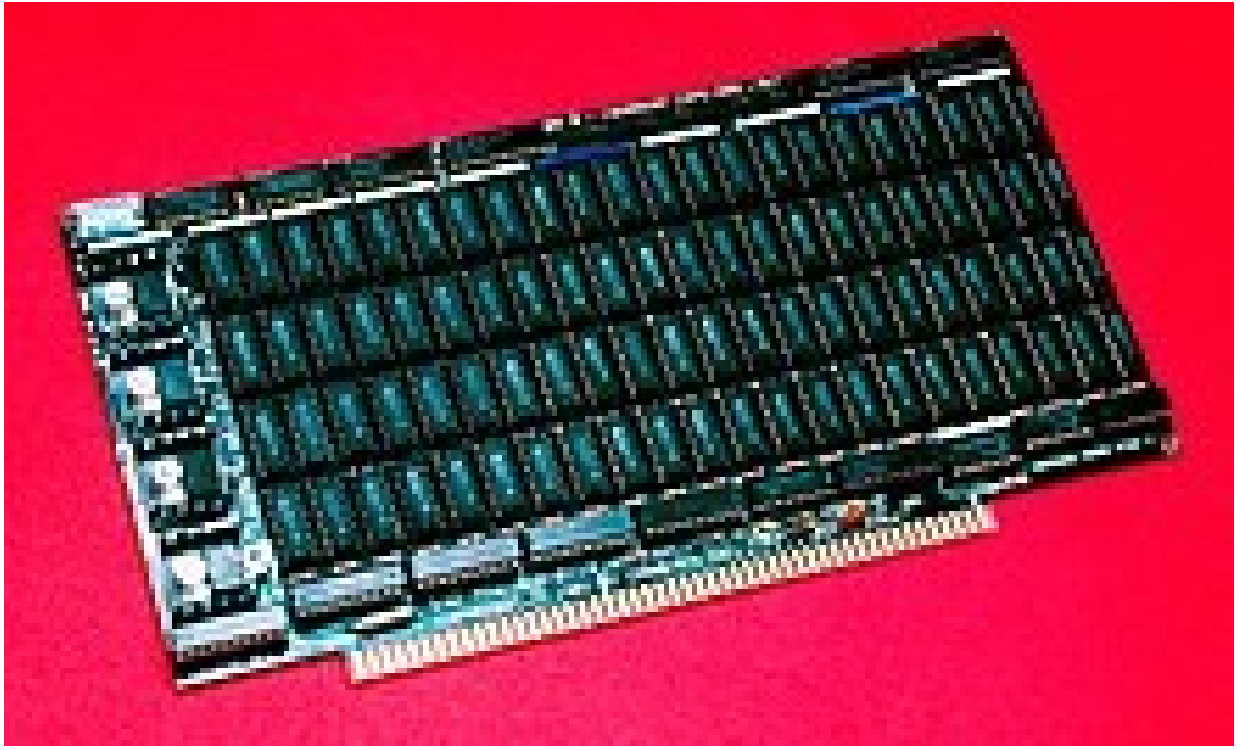
- An tòan: Đảm bảo hệ thống vận hành bình thường.
  - Khả năng chịu lỗi của hệ thống: là khả năng hệ thống vẫn hoạt động bình thường khi một số thành phần nào đó (nguồn, RAM, ổ cứng, vv..) của hệ thống bị hỏng hóc. Phương án: chạy song song (2 nguồn), Parity (RAM), RAID (ổ cứng), vv..
  - Khả năng phục hồi dữ liệu: Khi dữ liệu có vấn đề (bị mất mát, lỗi, phá hoại, ..) thì phải được phục hồi một phần hoặc hoàn toàn. Phương án: sao lưu định kỳ: hàng ngày (với dữ liệu quan trọng, thay đổi nhanh), hàng tuần hay hàng tháng (với dữ liệu ít thay đổi, không quá quan trọng).
  - Khả năng phục hồi hệ thống: sao lưu cả hệ thống, điện toán đám mây. Đảm bảo có các hệ thống cách nhau ít nhất 400km để đảm bảo có ít nhất một hệ thống vần hoạt động khi xảy ra các thiên tai lớn (lũ lụt, động đất,...)
- Bảo mật gồm 2 phần:
  - Đảm bảo dữ liệu không được tiết lộ cho bên thứ 3.
  - Đảm bảo tính toàn vẹn của dữ liệu dưới tác động của bên thứ 3.

Khả năng chịu lỗi của hệ thống

Trong quá trình vận hành, bất cứ một thiết bị điện tử nào cũng có khả năng hỏng hóc, vì vậy để tăng cường khả năng chịu lỗi, hầu hết các thành phần của các máy chủ đều gồm hai phần giống hệt nhau, và có khả năng thay thế trong lúc hoạt động (hot plug). Nguồn là một ví dụ. Bộ nguồn máy chủ không giống nguồn máy để bàn, nó có thể đơn giản rút ra mà không cần mở máy, để thay thế chỉ cần ấn bộ nguồn mới vào rãnh là được mà không cần cắm hay rút giây cắm nào khác ngoài dây nguồn. Ổ cứng cũng vậy, đơn giản chỉ là thao tác rút ra khỏi ray là được mà không cần phải tháo vit hay mở vỏ máy. Điều đó giúp nhân viên bảo trì dễ dàng thay thế thiết bị hỏng mà không cần phải tắt máy. Trong thời gian thiết bị bị rút ra khỏi máy, vẫn còn ít nhất một thiết bị khác đang hoạt động, giúp máy tính vẫn hoạt động bình thường.

Một số thành phần quan trọng khác như RAM hay ổ cứng có những phương thức khác, phức tạp hơn để đảm bảo dữ liệu người dùng an toàn.

Ram ECC - Error Correction Code



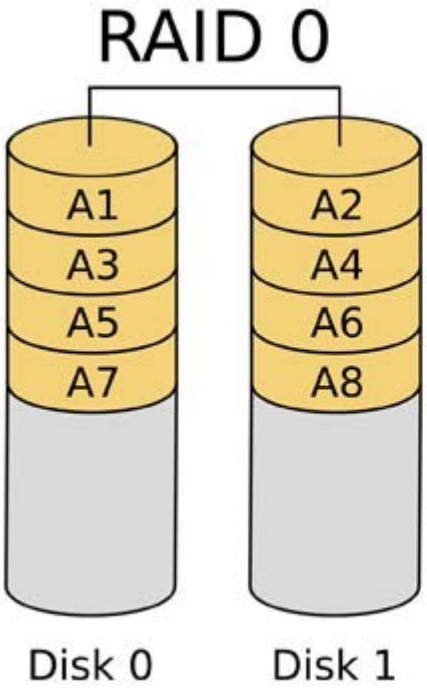
Thanh ram 512 KB dùng 22 bit lưu trữ
cho mỗi 2 byte - 6 byte dùng cho ECC

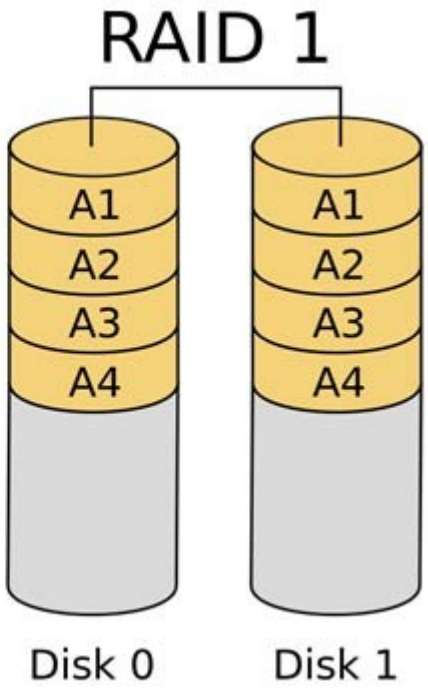
Ram DIMM ECC và
DIMM có parity.

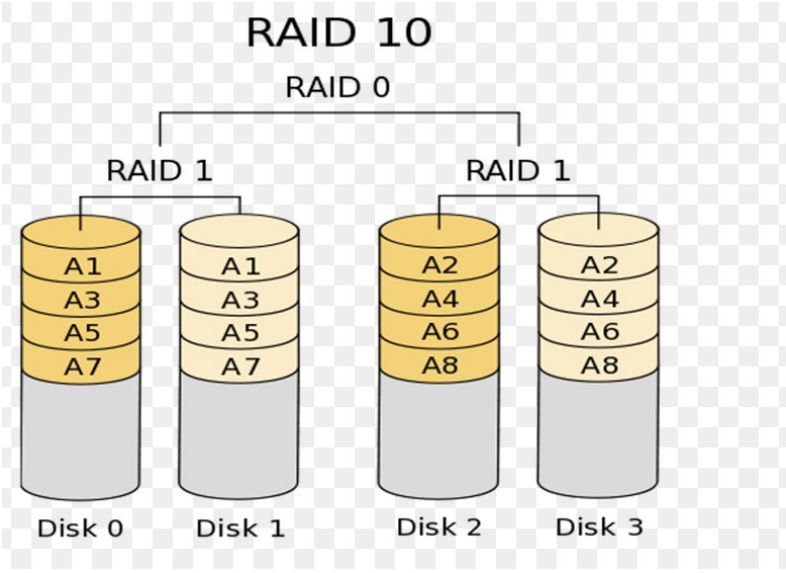
Ví dụ Error Correction

vị trí	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
giá trị	p1	p2	d1	p4	d2	d3	d4	p8	d5	d6	d7	d8	d9	d10	d11	p16
Parity bit	√		√		√		√		√		√		√		√	
		√	√			√	√			√	√			√	√	
				√	√	√	√					√	√	√	√	
								√	√	√	√	√	√	√	√	
																√

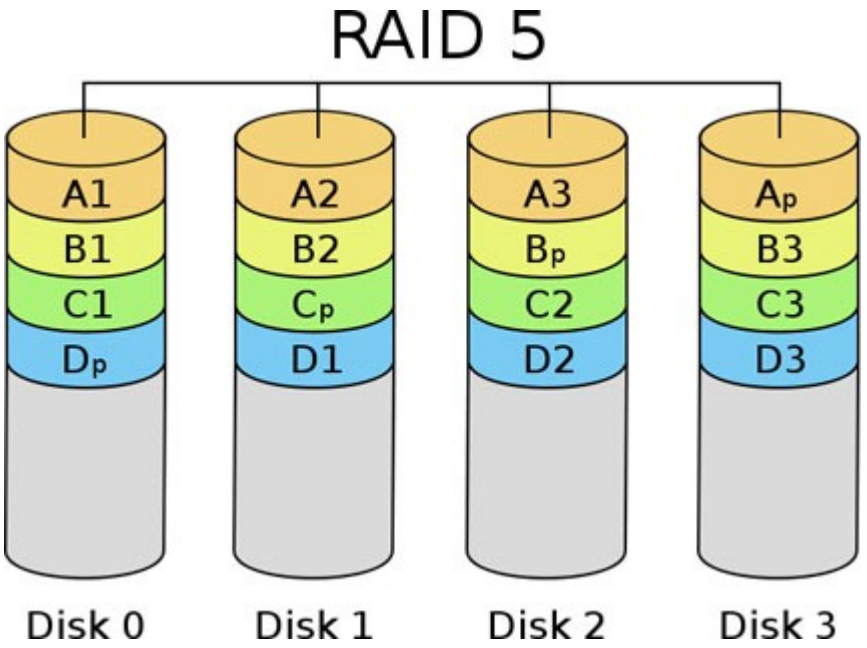
RAID – Redundant Arrays of Independent Disks

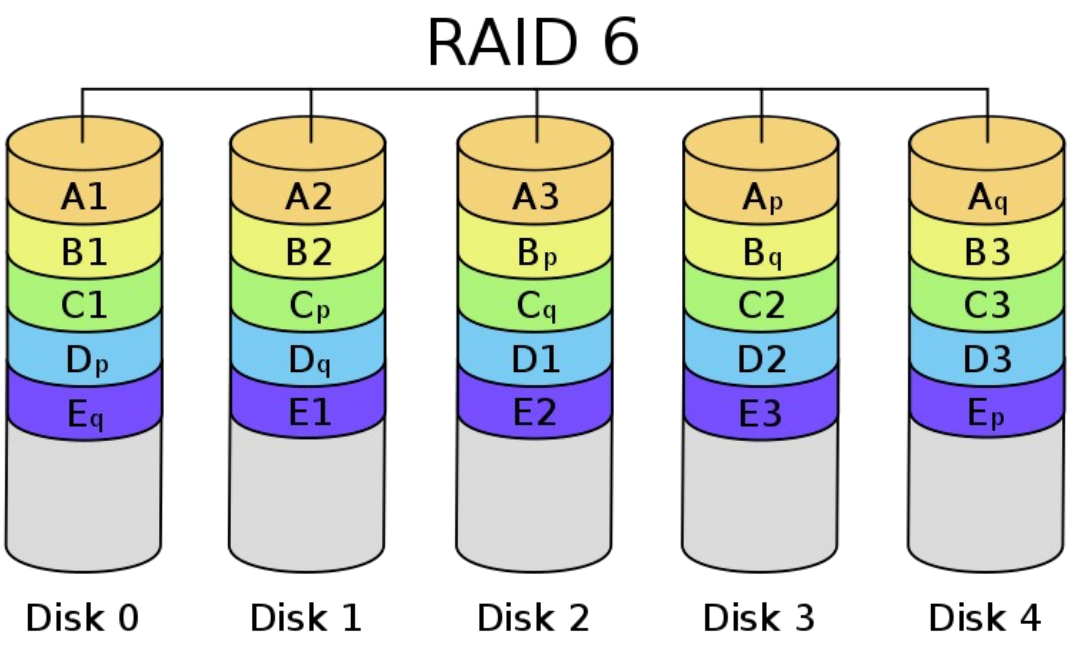






RAID – Redundant Arrays of Independent Disks





Thuật toán tính parity đơn giản nhất: Toán tử XOR

A_1 **xor** $A_2 = A_p <=> A_p$ **xor** $A_1 = A_2 <=> A_p$ **xor** $A_2 = A_1$

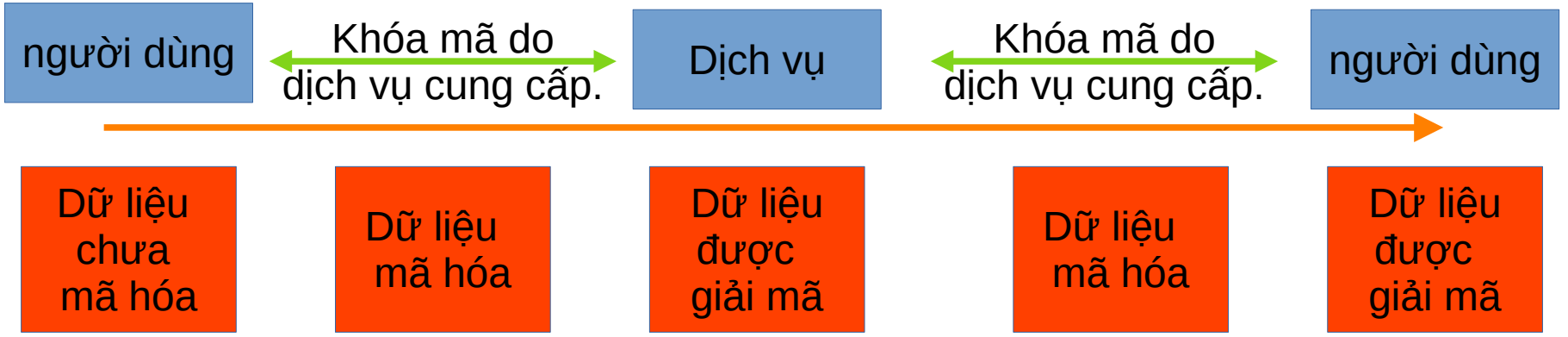
A_1 **xor** A_2 **xor** … **xor** $A_n = B <=> A_i = B$ **xor** A_1 **xor** .. **xor** A_{i-1} **xor** A_{i+1} **xor**.. **xor** A_n

Các lớp bảo mật

- Mã hóa đường truyền (TLS/SSL, E2E encrypt).
- Chứng thực danh tính công khai.
- Tường lửa (iptables/nftables)
- Hạn chế các lỗi nhập liệu (Tràn stack,...)
- Mật khẩu mạnh.

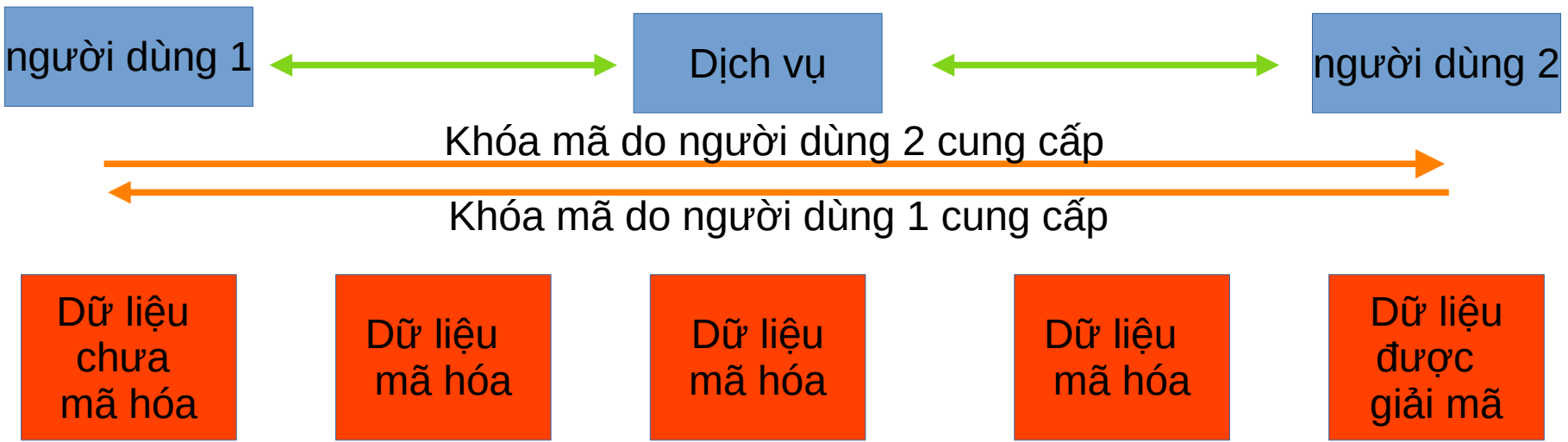
Mã hóa đường truyền

- TLS/SSL: Mã hóa bằng khóa công khai, khi kết nối với dịch vụ - do nhà cung cấp dịch vụ cung cấp. Thông thường là SHA 256 bit. Phương pháp này hạn chế nguy cơ nghe lén trên đường truyền. Nhưng nhà cung cấp dịch vụ có dữ liệu chưa mã hóa.

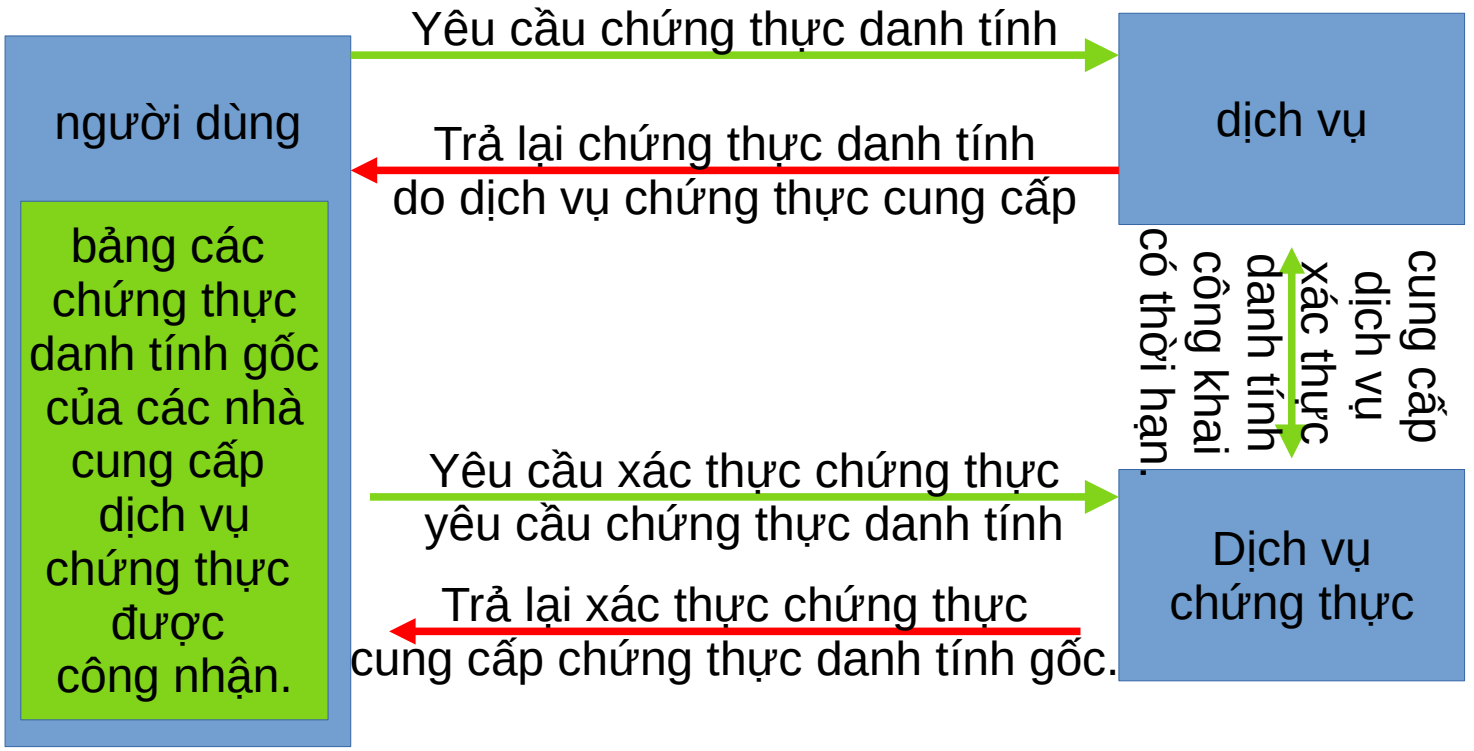


Mã hóa đường truyền

End to End encrypt: Mã hóa bằng khóa công khai do hai người dùng trao đổi với nhau. Vì vậy, chỉ có người dùng có dữ liệu chưa mã hóa. Có thể hoạt động trên đường truyền đã mã hóa.



Chứng thực danh tính



Tường lửa

- Chỉ mở các cổng cần thiết (các dịch vụ cho internet).
- Chỉ mở các IP cần thiết (IP của các máy có dịch vụ).
- Chỉ cho phép các IP hợp lệ.
- Khóa các IP login sai nhiều lần (3->5 lần) trong một khoảng thời gian ngắn nào đó (3->5 phút) để chống Brute-Force (tìm mật khẩu bằng phương pháp thử và sai theo từ điển)

Tường lửa

- Có rất nhiều ứng dụng tường lửa khác nhau. Hiện tại linux dùng NetFilter được tích hợp sẵn trong nhân, Người dùng có thể cùng nhiều phần mềm khác để quản lý, theo dõi NetFilter như IPTables, EBTables, IPRouter2,….
- BSD tích hợp sẵn PF (Packet Filter) module.
- NetBSD tích hợp NPF (Net Packet Filter) module.
- Windows được tích hợp sẵn Windows Firewall từ bản Windows XP SP2.

Mật khẩu mạnh

- Mật khẩu phải dễ nhớ, đảm bảo không cần dùng thiết bị hỗ trợ nhớ (sổ..).
- Mật khẩu không thể quá ngắn để đảm bảo an toàn, và không thể quá dài để tránh gõ sai, nên từ 8 đến 14 ký tự
- Đảm bảo có cả chữ HOA, chữ thường, số, ký tự đặc biệt (số giữ shift).
- Không liên quan đến các thông số mở của người dùng và những người thân rất gần (bố, mẹ, vợ, con, anh chị em ruột). Có thể dùng thông số mở của những người không quá liên quan, như bạn thân của bạn gái, bạn trai, bạn của anh chị ..
- Để đảm bảo các yêu cầu trên, nên dùng các sự kiện (đảm bảo có tên và ngày tháng) không đáng chú ý, nhưng liên quan trực tiếp đến bản thân (như các kỷ niệm không quan trọng liên quan đến sở thích - du lịch, gặp gỡ bạn, ...).
- Thay đổi thường xuyên (từ 6 tháng đến 1 năm).