**Phát hiện đăng nhập bất thường**

1. **Phát biểu bài toán**

Cho tập dữ liệu đăng nhập, xác định các sự kiện đăng nhập bất thường trong tập dữ liệu đầu vào.

* 1. **Mô tả dữ liệu**

Dữ liệu đầu vào gồm 1 tập các sự kiện đăng nhập, một sự kiện đăng nhập

đặc trưng bởi các thuộc tính sau:

- date: ngày đăng nhập

- account id: id của người đăng nhập

- src mask: dải ip chứa địa chỉ ip thực hiện việc đăng nhập

- target mask: dải ip chứa địa chỉ ip được đăng nhập

- src ip: địa chỉ ip thực hiện việc đăng nhập

- target ip: địa chỉ ip được đăng nhập

- src ip: địa chỉ ip thực hiện đăng nhập

- hour of day: thời điểm (giờ) đăng nhập trong ngày {0, 1,…,24}

- day of week: ngày đăng nhập trong tuần {Monday, Tuesday,…, Sunday}

- logon type: kiểu đăng nhập {2, 3, 4, 5, 7, 8, 9, 10, 11}

* 1. **Định nghĩa sự kiện đăng nhập bất thường**

**Định nghĩa 1:** Xét sự kiện đăng nhập là tập các kiện đăng nhập có cùng account id với sự kiện .

**Định nghĩa 2:** Xét sự kiện đăng nhập và thuộc tính là giá trị của thuộc tính trong sự kiện

**Định nghĩa 3:** Xét thuộc tính và thuộc tính v, là tập các sự kiện đăng nhập thỏa mãn:

**Định nghĩa 4:** Xét không gian , thuộc tính , là tập các sự kiện thỏa mãn:

**Định nghĩa 5:** Giả sử là tập các cụm thỏa mãn , gọi là 2 tham số đặc trưng cho ranh giới để xác định cụm bất thường nếu thỏa mãn 1 trong các điều kiện sau:

Trong đó, . Tập các cụm bất thường được xác định bởi .

Một sự kiện đăng nhập được coi là bất thường nếu thuộc một trong các trường hợp sau:

TH1:

* Coi một user được đặc trưng bởi cặp (account \_id, src\_ip), số lần user đăng nhập vào target ip là ít so với các target ip khác được coi là bất thường.
* Xét bộ 3 thuộc tính – ,

là tập các sự kiện thể hiện user đăng nhập vào target\_ip , trong đó . Chọn ra tập bất thường trên tập theo **định nghĩa 5**.

TH2:

* Coi một user được đặc trưng bởi cặp (account \_id, target\_ip), số lần user đăng nhập từ src\_ip là ít so với các src\_ip khác được coi là bất thường.
* Xét bộ 3 thuộc tính – ,

là tập các sự kiện thể hiện user đăng nhập từ src\_ip , trong đó . Chọn ra tập bất thường trên tập theo **định nghĩa 5**.

TH3:

* Coi một user được đặc trưng bởi bộ 3 (account \_id,src\_ip, target\_ip), số lần user đăng nhập bằng logon\_type là ít so với các logon\_type khác được coi là bất thường.
* Xét bộ 4 thuộc tính – , là tập các sự kiện thể hiện user đăng nhập bằng logon\_type , trong đó . Chọn ra tập bất thường trên tập theo **định nghĩa 5**.