U, *V* nhận được mã số nhân dạng sau:

$ID(U) = h(f(U)) = 87,954,543$

$ID(V) = h(f(V)) = 758,355,475$

Các mã số này sau đó được sử dụng làm giá trị tìm kiếm theo trường khóa chính Checksum trong CSDL luật của hệ.

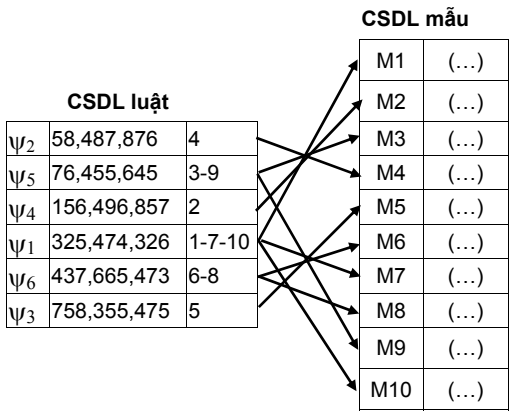
3.2.3 Truy vấn luật

Khi tra cứu mã số nhân dạng *ID*(*A*) trong CSDL luật, nếu không tìm thấy có thể kết luận *A* là đối tượng an toàn vì trong CSTT chưa có các mô tả mã độc nào có đặc trưng thi hành giống với *A*. Nếu tìm thấy, cần đưa *A* vào diện nghi vấn, sau đó tiếp tục xác minh các đặc điểm nhân dạng đặc thù để có kết luận chính xác. Trong ví dụ trên, *U* là an toàn vì trong tập luật không có giá trị checksum 87,954,543 nào. Tuy nhiên cần cảnh giác *V* vì mã số nhân dạng 758,355,475 của nó đang có trong CSDL luật.

Mỗi luật nhận dạng đại diện cho một cụm (nhóm) các mã độc cùng đặc trưng. Xác minh hồ sơ là quá trình đối chiếu thông tin nhân dạng của *A* với các mã độc cùng nhóm. Cho các số nguyên *i*, *j*; hàm băm *H* truy vấn danh sách mã độc cùng nhóm trên tập mẫu *X(p,k)* có dạng:

$$H(\psi_m) = \psi_i \gg \{X(p, j)\} \ \forall \ (i < m); (j < k)$$

Theo thiết kế này, hàm *H* chỉ đơn giản trả về danh sách ở trường thứ ba của nút lá trên nhánh V-Tree của luật ψ_i tương ứng. Chiến lược sử dụng tập luật nhận dạng làm chỉ mục băm tìm kiếm chữ ký mã độc được minh họa ở Hình 3.



Hình 3: Cơ chế băm tập luật theo chỉ mục

3.2.4 So khớp mẫu

Đây là công đoạn rà soát, đối chiếu thông tin lần cuối trước khi quyết định 'bắt giam' đối tượng nhằm hạn chế sai sót trong tác nghiệp. Có nhiều cách so khớp mẫu: dò tìm mã phổ biến, trích xuất các giá trị đặc trưng tại các địa chỉ nhạy cảm, trích chọn mã ngẫu nhiên…[5].

Trong bài viết này, so khớp mẫu là bước kiểm tra sự xuất hiện giá trị thuộc tính thứ *p* của quan hệ *X*(*p*,*k*) trong đối tượng chẩn đoán. Nếu đối chiếu đúng, có thể kết luận đối tượng *A* là mã độc. Ngược lại, tiếp tục đối chiếu các mẫu còn lại cho đến khi phát hiện. Gọi *M*(*p*,*n*) là tập mã độc có mã nhận dạng trùng với mã nhân dạng của *A*. Giải thuật so khớp mẫu như sau:

```
Tìm thấy ← false
j ← 1
Lặp lại
    Nếu Chữ ký(A) = M(p,j) thì
       Kết luận ('Phát hiện mã độc', M(1,j))
       Tìm thấy ← true;
    Ngược lại
       j ← j+1
Đến khi Tìm thấy hoặc (j>n)
```

4 KẾT QUẢ THỰC NGHIỆM

4.1 Thử nghiệm với D2 Anti-virus*

*4.1.1 Giới thiệu hệ D2 Anti-virus**

D2 Anti-virus* (Diagnose and Destroy Computer Viruses) là phần mềm quét virus hướng tiếp cận máy học của Việt Nam. Thiết kế theo mô hình của một hệ chuyên gia, hoạt động của hệ D2 gồm ba giai đoạn: giao tiếp chuyên gia (giai đoạn học), động cơ suy diễn (giai đoạn nhận dạng, xử lý) và giao tiếp người dùng (giai đoạn tổng kết, báo cáo).

Thực hiện trên máy chuyên gia, giai đoạn học xây dựng CSTT dưới dạng tập các mô tả tri thức mã độc và luật nhận dạng khẳng định dương trên tập mẫu. Giai đoạn xử lý và báo cáo thực hiện trên máy khách. Hoạt động hướng tác tử [10], động cơ quét của hệ chứa các thuật toán học phân loại, động cơ suy diễn, thực thi các lớp bài toán học và xử lý các trường hợp lây nhiễm. Cuối cùng, giai đoạn báo cáo tổng hợp kết quả các bài toán, phân tích quá trình, tạo các giao tiếp hội thoại, thu nhận ý kiến người dùng, báo cáo kết quả và tăng trưởng tri thức.

Phiên bản D2 Anti-virus* 2013 (Hình 4) có một số cải tiến quan trọng về giao diện, tăng tốc độ quét, dự báo mã độc hướng heuristic, ước lượng mã tương đồng, phát hiện hành vi lây nhiễm trên thiết bị lưu trữ cá nhân…