**1.Định nghĩa và một số tính chất của dàn**

Trước hết ta nhắc lại rằng với , không gian là không gian định chuẩn trên với chuẩn của phần tử được định nghĩa bởi:

.

**1.1 Định nghĩa.**Một tập hợp khác rỗng được gọi là một *dàn* nếu là một *nhóm cọng con rời rạc* của .

Ở đây tính chất rời rạc có nghĩa là tồn tại số sao cho với mọi thì .

Chú ý rằng định nghĩa trên không cho ta một hình ảnh hình học về dàn. Sau đây ta sẽ đưa ra một định nghĩa tương đương về dàn, định nghĩa này cho phép hình dung rõ ràng hơn về mặt hình học.

Giả sử  là các vector trong . Ta kí hiệu  là tập gồm các tổ hợp tuyến tính của  với hệ số nguyên, tức là

.

Rõ ràng  là một nhóm cọng con của , tuy nhiên không nhất thiết có tính chất rời rạc. Chẳng hạn khi , tập  là không rời rạc. Nếu là một dàn, ta bảo rằng sinh bởi hay  là các phần tử sinh ra dàn . Nếu thêm  là độc lập tuyến tính, ta bảo hệ là một cơ sở của , như vậy mỗi vector trong dàn sẽ biểu diễn dưới dạng một tổ hợp tuyến tính với hệ số nguyên qua các vector trong cơ sở.

Số lớn nhất các vector độc lập tuyến tính trong một dàn được gọi là *số* *chiều* (dimension) hoặc *hạng* (rank) của và được kí hiệu bởi . Nếu  là một dàn với thì ta bảo có chiều đầy đủ. Trong trường hợp thì rõ ràng trong luôn tồn tại vector độc lập tuyến tính, nhưng các vector này có thể không tạo thành một cơ sở của . Tuy nhiên định lý sau đây chỉ ra rằng ta luôn xây dựng được một cơ sở của từ các vector này.

**1.2 Định lý.** Giả sử là một dàn trong ,  và  là các vector độc lập tuyến tính trong . Khi đó sẽ tồn tại ma trận tam giác dưới sao cho các vector  xác định bởi tạo thành một cơ sở của .

*Chứng minh.* Với , xét tập hợp .

Tập hợp này là hữu hạn do nếu  thì và tập  là hữu hạn, ngoài ra do . Do đó tồn tại phần tử dương bé nhất của , kí hiệu là . Theo định nghĩa , sẽ tồn tại các phần tử sao cho . Ta sẽ chứng minh rằng  là một cơ sở của .

Do nên hệ  là độc lập tuyến tính. Với thì do hệ  là hệ độc lập tuyến tính tối đại, sẽ tồn tại các số sao cho . Kí hiệu  và đặt , ta có  và . Nếu tất cả các  không nguyên, ta kí hiệu  là chỉ số lớn nhất mà , khi đó và  nếu . Từ



do ta có mâu thuẩn với  là phần tử bé nhất của . ■

Bây giờ giả sử  là một cơ sở của dàn L và  là một cơ sở khác. Do  là cơ sở nên mỗi vector  có thể biểu diễn qua cơ sở này:

.

Ma trận chuyển đổi cơ sở từ  sang  là . Khi đó ma trận chuyển cơ sở ngược lại là ma trận . Do các phần tử của  và  là các số nguyên nên  là các số nguyên. Từ tính chất  ta có ngay . Ngược lại nếu  thì ma trận  sẽ có các phần tử là các số nguyên và do đó  là một cơ sở của L. Điều này chứng tỏ:

**1.3 Định lý.** Ma trận chuyển đổi giữa hai cơ sở của dàn có các phần tử nguyên và định thức bằng .

Như vậy, một dàn bất kỳ trong  đều tồn tại một cơ sở. Cho dù tập có dạng  là một dàn hay không, một dàn bất kỳ đều có dạng với  là hệ độc lập tuyến tính. Điều ngược lại sau đây có thể chứng minh dễ dàng.

**1.4 Định lý.** Nếu  là hệ độc lập tuyến tính trong thì tập  là một dàn với .

Như không gian vector, một dàn có nhiều cơ sở. Phần tiếp theo sẽ khảo sát một khái niệm mới, bất biến qua cơ sở của dàn.

**1.5 Định nghĩa.** Giả sử *L* là một dàn với số chiều *k* và là một cơ sở của *L*. *Miền cơ bản* (hay *hình bình hành cơ bản*) *của L*, được định nghĩa là tập



Định lý sau đây cho biết vì sao miền cơ bản đóng vai trò quan trọng trong việc khảo sát dàn.

**1.6Mệnh đề.** Giả sử là một dàn với số chiều *n* và *F* là miền cơ bản của *L*. Khi đó mỗi vector sẽ có biểu diễn duy nhất dạng  với  và .

*Chứng minh.* Giả sử  là cơ sở của L và . Khi đó  cũng là một cơ sở của . Do đó vector sẽ có biểu diễn  với . Bằng cách viết  dưới dạng  với  và , ta có  với  và . Bây giờ giả sử với  và . Nếu  và  thì . Do  là độc lập tuyến tính, ta có ngay  hay . Vì  và nên . Vậy và ■

Như vậy miền cơ bản sẽ phụ thuộc vào cơ sở của dàn. Phần tiếp theo giới thiệu một đặc trưng của miền cơ bản không phụ thuộc vào cơ sở.

**1.7 Định nghĩa.** Giả sử *L* là một dàn với số chiều bằng *n* và *F* là miền cơ bản của *L*. Độ đo hay thể tích của *F* là  sẽ được gọi là *định thức của dàn L* và kí hiệu là *det(L).*

Trong trường hợp *n=2*, nếu 2 vector cơ sở của dàn *L* vuông góc nhau thì ta có . Tổng quát hơn ta có bất đẳng thức sau.

**1.8 Mệnh đề.(bất đẳng thức Hadamard)** Nếu L là một dàn với cơ sở  và F là miền cơ bản thì .

Mệnh đề sau đây cho cách tính định thức của dàn L theo các vector cơ sở của nó.

**1.9 Mệnh đề.** Giả sử  là dàn với số chiều *n* và  là miền cơ bản của *L* ứng với cơ sở  . Khi đó nếu thì định thức của *L* tính bởi công thức  trong đó.

*Chứng minh.* Độ đo hay thể tích của  trong  được tính bởi

.

Thực hiện phép đổi biến số từ các biến  sang các biến  với . Rõ ràng ma trận Jacobian của phép đổi biến số trên là  và miền cơ bản F là ảnh của hình hộp đơn vị  qua phép đổi biến, do đó .■

Một hệ quả của mệnh đề trên là định thức của dàn không phụ thuộc vào cơ sở của dàn, được khẳng định bởi mệnh đề sau.

**1.10 Hệ quả.** Giả sử  là một dàn có số chiều n. Khi đó mọi miền cơ bản của L đều có cùng thể tích. Do đó định thức  là một bất biến của dàn, không phụ thuộc vào cơ sở.

*Chứng minh.* Giả sử  và  là hai cơ sở của *L*, và là các miền cơ bản của L tương ứng với 2 cơ sở này và ,  là các định thức tương ứng (định nghĩa ở mệnh đề 1.9). Nếu kí hiệu A là ma trận chuyển cơ sở từ  sang  thì

.

Do đó

 





 (do *|det(A)|=1* theo mệnh đề 1.3)

. ■

**2. Một số bài toán trên dàn.**

Cho  là một dàn, .

**2.1 Bài toán tìm vector ngắn nhất (SVP-The Shortest Vector Problem).**

Tìm  mà .

**2.2 Bài toán tìm vector gần nhất (CVP-The Closest Vector Problem).**

cho trước , tìm mà .

**2.3 Bài toán tìm cơ sở ngắn nhất (SBP-The Shortest Basis Problem).**

tìm cơ sở  của L ngắn nhất theo một nghĩa nào đó, chẳng hạn  hay .

**2.4 Bài toán tìm vector xấp xỉ ngắn nhất:**

cho  là một hàm theo *n*, tìm  mà .

Chú ý rằng trong một số trường hợp, việc giải các bài toán trên là rất đơn giản. Chẳng hạn giả sử L có một cơ sở trực giao là .

Bài toán SVP: do vector sẽ có biểu diễn nên . Từ đây ta có ngay vector có độ dài ngắn nhất trong các vector  chính là lời giải của bài toán SVP.

Bài toán CVP: với  và  ta có .

Từ đây, lời giải của bài toán CVP là vector *v* tương ứng với các hệ số là các số nguyên gần  nhất.

**3. Tìm cơ sở thu gọn của dàn.**

Phần này trình bày 2 thuật toán tìm cơ sở thu gọn của một dàn cho trước. Đó là thuật toán Gauss trong trường hợp 2 chiều và tổng quát là thuật toán LLL. Các thuật toán này đóng vai trò quan trọng trong việc giải các bài toán ở phần 2 và áp dụng nhiều trong lĩnh vực thám mã.

**3.1 Thuật toán Gauss.**

Giả sử là một dàn,  và  là một cơ sở đã biết của *L*. Thuật toán Gauss cho phép từ cơ sở  xây dựng một cơ sở khác có độ dài ngắn hơn. Ý tưởng của Gauss là thay  bằng  với . Tuy nhiên do  có thể không nguyên nên có thể , do đó ta thay  với , trong đó  chỉ số nguyên gần *x* nhất.

**3.1.1Mệnh đề.** Nếu thực hiện theo thuật toán sau:

***Loop***

***If***  ***then*** hoán đổi .

Tính .

***If*** m=0 ***then*** STOP ***Else*** .

***EndLoop***.

Ta sẽ thu được cơ sở mới với là vector ngắn nhất của *L*. Hơn nữa nếu  là góc giữa  thì  và .

*Chứng minh.* Khi thuật toán kết thúc ta phải có  và  (\*).

Một vector bất kỳ sẽ có dạng , khi đó:

 





 (do (\*))

 ( do )



 (chú ý  do  ).

Ngoài ra khi thuật toán kết thúc ta có , do đó nếu là góc giữa và thì . Vậy hay .■

**3.1.2 Ví dụ.** Xét dàn 2 chiều trong với cơ sở gồm 2 vector . Bằng cách thực hiện thuật toán Gauss trên cơ sở này, ta có kết quả:

|  |  |  |  |
| --- | --- | --- | --- |
| **Bước lặp** | **V1** | **V2** | **m** |
| **1** | **(6513996,6393464)** | **(66586820,65354729)** | **10** |
| **2** | **(1446860,1420089)** | **(6513996, 6393464)** | **5** |
| **3** | **(-7203041,-706981)** | **(1446860,1420089)** | **-2** |
| **4** | **(6252,6127)** | **(-7203041,-706981)** | **-115** |
| **5** | **(-1324,-2376)** | **(6252,6127)** | **-3** |
| **6** | **(2280,-1001)** | **(-1324,-2376)** | **0** |

Ta nhận được một lời giải cho bài toán SVP là (2280,-1001).

**3.2 Thuật toán LLL.**

Trước hết chúng ta nhắc lại một số khái niệm cần thiết cho việc giới thiệu thuật toán Gram-Schmidt.

**3.2.1Thuật toán trực giao hóa Gram-Schmidt.**

Thuật toán trực giao hóa Gram-Schmidt nhằm xây dựng từ một cơ sở  của không gian vector một cơ sở mới thỏa mãn:

i.Cơ sở  là cơ sở trực giao, tức là .

ii. .

Thuật toán Gram-Schmidt như sau:

Đặt .

**For** i=2 **to** m **do**

Tính .

Đặt .

**Endfor**

Bây giờ nếu  là một cơ sở của dàn thì có thể  không là cơ sở của *L*. Tuy nhiên ta có mối liên hệ sau.

**3.2.2Mệnh đề.** Nếu  là một cơ sở của dàn  và là cơ sở trực giao nhận được từ cơ sở  bằng thuật toán Gram-Schmidt. Khi đó .

*Chứng minh.* Dùng kí hiệu ,  như ở phần 1.5. Theo thuật toán Gram-Schmidt ta có  với *M* là ma trận chuyển cơ sở:

.

Do nên .■

**3.2.3 Chú ý.**Trong lý thuyết không gian vector, ta đã biết rằng nếu  là không gian vector, W là không gian vector con của V và là không gian con trực giao với W thì  cũng là không gian vector con của V và , tức là mỗi vector  có biểu diễn duy nhất dưới dạng  với . Khi đó trong thuật toán Gram-Schmidt,  là hình chiếu của lên .

Nếu  là cơ sở của dàn  và là cơ sở trực giao nhận được từ cơ sở bằng thuật toán Gram-Schmidt thì có thể  không là cơ sở của L, tuy nhiên khái niệm cơ sở LLL của dàn L lại được định nghĩa thông qua .

**3.2.4 Định nghĩa.** *Giả sử  là cơ sở của dàn  và là hệ nhận được từ bằng thuật toán Gram-Schmidt. Khi đó được gọi là một LLL-cơ sở thu gọn của L nếu và chỉ nếu:*

*i.(Điều kiện về kích thước) .*

*ii.(Điều kiện Lovász).*

Rõ ràng: Điều kiện Lovász 

h/c của lên h/c của lên .

Một công trình của Lenstra, Lenstra và Lovász chỉ ra rằng LLL-cơ sở thu gọn là một cơ sở tốt và có thể tìm được một LLL-cơ sở thu gọn từ một cơ sở đã biết trong thời gian đa thức. Mệnh đề sau chỉ ra rằng LLL-cơ sở có tính chất mong muốn.

**3.2.5 Mệnh đề.***Giả sử L là dàn với số chiều n. Một LLL-cơ sở thu gọn  của L sẽ có 2 tính chất sau:*

*i.*

*ii.Vector đầu tiên  thỏa mãn  và  .*

*Như vậy, một LLL-cơ sở thu gọn chính là lời giải của bài toán apprSVP với hệ số .*

*Chứng minh.* Điều kiện Lovász và điều kiện  suy ra rằng

 (1) .

Áp dụng bất đẳng thức (1) một số lần ta thu được:

 (2).

Do đó

 

 (do hệ  là hệ trực giao)

 (do )

 ( do (2))

 (3).

Nhân các bất đẳng thức (3) khi  ta có:

.

Lấy căn bậc hai bất đẳng thức trên ta nhận được điều đầu tiên trong Mệnh đề 4.5.i.

Tiếp theo với , từ (3) ta có

.

Lấy căn bậc hai bất đẳng thức trên ta nhận được điều còn lại trong Mệnh đề 4.5.i.

Trong bất đẳng thức sau của 4.5.i, lấy  và nhân các bất đẳng thức khi  ta có

.

Lấy căn bậc n hai vế ta có điều đầu tiên trong 4.5.ii.

Để chứng minh phần còn lại trong 4.5.ii, giả sử là vector khác 0 và có biểu diễn

 , trong đó , các số là các số nguyên trong khi các số là các số thực, ngoài ra . Theo thuật toán Gram-Schmidt, với , các vector ** là đôi một trực giao và  nên

 và . Từ đây ta có , do đó . Kết hợp điều này với 4.5.i cho ta

.

Lấy căn bậc hai 2 vế, ta nhận được bắt đẳng thức còn lại trong 4.5.ii.■

**3.2.6 Thuật toán tìm LLL-cơ sở thu gọn.**

**Input**: cơ sở **của dàn L.

Đặt .

**Loop while** 

**For** j=1 **to** k-1

Đặt 

**Endfor**

**If**   **then**

Đặt k=k+1

**Else**

Hoán đổi  và 

Đặt 

**Endif**

**EndLoop**

**Output**: LLL-cơ sở thu gọn **.

Tại mỗi bước lặp, **là cơ sở trực giao thu được từ cơ sở **bằng thuật toán Gram-Schmidt và .

Mệnh đề sau cho biết tính hiệu quả của thuật toán trên.

**3.2.7Mệnh đề.** Giả sử là cơ sở của dàn *L*. Thuật toán 4.6 sẽ chấm dứt sau một số hữu hạn bước và trả về một LLL-cơ sở thu gọn của *L*. Một cách cụ thể, nếu  thì vòng lặp Loop trong thuật toán trên sẽ được thực hiện không quá  lần. Như vậy, thuật toán LLL trên là thuật toán với thời gian đa thức.

Chứng minh. Có thể tham khảo trong [1].

**3.2.8Chú ý.** Một trong những tiêu chuẩn để đánh giá cơ sở này *tốt hơn* cơ sở khác là dựa vào tỷ số Hadamard. Chúng ta đã biết bất đẳng thức Hadamard

.

Trong đó ** là cơ sở của dàn L, dấu bằng xảy ra khi các vector trong cơ sở đôi một trực giao. Tỷ số Hadamard của cơ sở ** được định nghĩa là

.

Rõ ràng  với mọi cơ sở B, nếu giá trị *H(B)* càng gần 1, cơ sở *B* càng tốt.

**3.2.9 Ví dụ.**Xét dàn L sinh bởi cơ sở *B* gồm 6 vector dòng của ma trận

.

Bằng cách dùng thuật toán LLL, ta thu được LLL cơ sở  thu gọn là các hàng của ma trận

.

Có thể tính được  và . Điều này cho thấy cơ sở *B* là tốt hơn cơ sở .

**4.Áp dụng vào thám mã.**

**4.1 Tấn công RSA trong trường hợp khóa yếu**

(Public Key Cryptanalysis, Phong Q. Nguyen, Recent Trends in Cryptography, 2008)

Xét hệ mã hóa RSA trong trường hợp  và các số là cùng chiều dài bit. Ta có các ước lượng sau:

.

Do  nên .

Như vậy, nếu đặt  thì .

Xét dàn 2 chiều trong  với 2 vector cơ sở là . L sẽ chứa vector  mà , ngoài ra . Do  nên . Ta dự đoán  là vector ngắn nhất của *L*. Bằng cách dùng thuật toán Gauss cho cơ sở khởi đầu  ta tìm ra *t*, từ đó tìm ra *d*.

**Ví dụ.** Chọn . Khi đó . Chọn (điều kiện  được thỏa mãn).

Từ khóa chung (e,n), kẻ tấn công sẽ tính được và xuất phát từ cơ sở 



bằng cách thực hiện thuật toán Gauss, kẻ tấn công nhận được cơ sở thu gọn



.

Từ đây kẻ tấn công sẽ tính được  .

**4.2 Thám mã hệ mã hóa Merkle-Hellman.**

**4.2.1 Bài toán tổng các tập con(subset-sum problem)**

Cho một danh sách các số nguyên dương (M1,M2,…,Mn) và một số nguyên dương S. Hãy tìm tập con trong danh sách có tổng các phần tử bằng S. Ta sẽ kí hiệu ngắn gọn bài toán này là (M,S).

Như vậy bài toán đưa về tìm chuỗi nhị phân x=(x1,x2,…,xn) sao cho . Thay vì kiểm tra vét cạn trên 2n chuỗi nhị phân độ dài n, ta có thuật toán sau.

**4.2.2 Mệnh đề.** Giả sử M=(M1,M2,…,Mn) và (M,S) là bài toán tổng các tập con. Sẽ tồn tại các tập con và sao cho , trong đó

và .

Tuy nhiên bài toán vẫn có thể có nhiều lời giải. Dưới đây là điều kiện đảm bảo tính duy nhất của lời giải.

**4.2.3 Định nghĩa.** Một *dãy siêu* tăng (superincreasing sequence) các số nguyên là một dãy các số nguyên dương

thỏa mãn tính chất .

Có thể kiểm tra được rằng nếu là một dãy siêu tăng các số nguyên thì

.

**4.2.4 Mệnh đề.** Xét bài toán (M,S) với các số nguyên trong S lập thành một dãy siêu tăng. Khi đó nếu lời giải bài toán tồn tại thì nó là duy nhất và có thể được tìm bằng thuật toán nhanh như sau:

**For** i **from** n **down to** 1 **do**

**If** **then** =1, S=S-Mi

**Else** xi=0

**Endfor**

*Chứng minh.* Mệnh đề là hệ quả trực tiếp của tính chất dãy siêu tăng.

**4.2.5 Hệ mã hóa Merkle-Hellman.**

Sinh khóa. -Chọn dãy siêu tăng .

-Chọn hai số nguyên A,B thỏa B>2rn và gcd(A,B)=1.

-Tính Mi=Ari(mod B) với .

-Công bố khóa chung M=(M1,M2,…,Mn).

Mã hóa. -Văn bản là các chuỗi nhị phân x có độ dài n.

-Tính S=x.M.

Giải mã. -Tính S’=A-1S(mod B).

-Giải bài toán (r,S’) được lời giải là x, tức x.r=S’.

**4.2.6 Chú ý.** Giả sử đã chọn giá trị n. nếu r1 nhỏ thì hệ mã rất dễ bị tấn công, do đó ta chọn r1>2n . Khi đó r2>2n+1,…,rn>22n, B>2rn>22n+1. Như vậy Mi=0(22n) và S=0(22n).

**4.2.7 Ví dụ.** Chọn dãy siêu tăng r=(3,11,24,50,115), A=113, B=250.

Tính được M=(113.3,113.11,113.24,113.50,113.115)(mod 250)=(89,243,212,150,245).

Văn bản m=(1,0,1,0,1) được mã hóa thành S=x.M=1.89+0.243+1.212+0.150+1.245=546.

Giải mã như sau: Tính 113-1(mod 250)=177.

Tính S’=A-1S(mod B)=177.546(mod 250)=142.

Giải bài toán tổng tập con S’=x.r, giải ra x=(1,0,1,0,1).

**4.2.8 Tấn công hệ mã Merkle-Hellman bằng dàn.**

-Giả sử lấy được văn bản đã mã hóa S là tống của bài toán tồng tập con (S,M) với M=(m1,m2,…,mn).

-Lập ma trận

.

Các vector hàng của ma trận trên là v1,v2,…,vn,vn+1.

-Xét dàn L={a1v1+…+anvn+an+1vn+1:a1,…,an,an+1 nguyên}.

-Giả sử x=(x1,x2,…,xn) là lời giải của bài toán tổng tập con. Khi đó L phải chứa vector

.

Do xi bằng 0 hoặc 1 nên . Mặt khác, mi=0(22n) và S=0(22n), tức là các vector sinh ra L có độ dài ||vi||=0(22n). Dường như *t* là vector khác 0 nhỏ nhất trong *L*. Bằng cách áp dụng thuật toán LLL cho các vector dòng của *L*, ta tìm được vector “ngắn” của *L*, tư đó tìm ra *xi.*

**4.3 Thám mã hệ mã hóa NTRU.**

Với các số nguyên dương N,p,q, d1, d2 ta kí hiệu:.

*.*

-Phép nhân hai phần tử trong R chính là phép nhân hai đa thức, tức là nếu

a(x)=a0+a1x+…+aN-1xN-1

b(x)=b0+b1x+…+bN-1xN-1

thì

c(x)=a(x).b(x)=c0+c1x+…+cN-1xN-1 với .

-Với , phần tử trung tâm tương ứng với a(x) trong R là đa thức (duy nhất) sao cho và các hệ số của a’(x) ở trong khoảng .

**4.3.1 Hệ mã hóa NTRU.**

Chọn các số nguyên dương N,p,q,d với N,p nguyên tố, gcd(p,q)=gcd(N,q)=1 và q>(6d+1)p.

Sinh khóa. -Chọn là khả nghịch trong Rp và Rq.

-Chọn và chọn ngẫu nhiên .

-Tính trong Rq và trong Rp.

-Khóa chung là và *r*.

Mã hóa. -Văn bản được mã hóa bằng cách tính .

Giải mã . -Tính . a(x)=b’(x) là phần tử trung tâm tương ứng với b(x).

-Tính

**4.3.2 Ví dụ.** (N,p,q,d)=(7,3,41,2).

f(x)=x6-x4+x3+x2-1, g(x)=x6+x4-x2-x, r(x)=x6-x5+x-1.

+31.

+2.

.

Khóa chung: (h,r).

Mã hóa: văn bản được mã hóa thành

*.*

Giải mã:Tính

.

.

Phần tử trung tâm ứng với kết quả trên là .

**4.3.3 Thám mã hệ mã hóa NTRU bằng dàn.**

Giả sử . Xét dàn số chiều 2N sinh bởi các hàng của ma trận với I là ma trận đơn vị và

.

Người ta chứng minh được các kết quả sau.

-, trong đó vector *(f,g)* gồm các hệ số của *f* và *g* viết theo thứ tự từ thấp đến cao.

-Với và thì:

..

.Vector ngắn nhất của có độ dài .

-Như vậy khi N rất lớn, khả năng vector ngắn nhất của là (f,g).