**1 MÃ HÓA MẬT MÃ**

**MẬT MÃ = MÃ HÓA + THÁM MÃ**

**Khái niệm**

• **Mật mã học** – **cryptology**. Nghiên cứu về mã hóa và thám mã. Mật mã học một ngành của toán học nghiên cứu lý thuyết số và áp dụng các công thức và thuật toán làm nền tảng cho mã hóa thông tin và phân tích các thông tin được mã, gọi là thám mã.

• **Mã hóa – cryptography**. Nghệ thuật và khoa học tạo ra mật mã. Mã hóa là quá trình viết bằng nhiều cách "mật mã" khác nhau để giữ bí mật thông điệp. Điều này liên quan đến việc thiết kế các hệ mã. Tài liệu này chỉ tập trung vào một số khái niệm toán học then chốt của mật mã.

• **Thám mã – cryptanalysis**. Nghệ thuật và khoa học phá mã. Là việc trích xuất các thông tin từ hệ mã và tiến trình mã hóa nhằm mục đích nghiên cứu liên quan phá vỡ các hệ mã.

• **Hệ mã – cryptosystem**. Cài đặt các thuật toán và giao thức mã hóa, giải mã và tạo khóa.

• **Hệ thống mật mã – cryptographic system**. Là bất kỳ hệ thống nào sử dụng mật mã.

• **Mật mã – cipher**. Thuật toán được sử dụng trong một hệ mã.

• **Độ hỗn loạn – confusion**. Tính chất của mối quan hệ giữa thông điệp, bản mã và khóa phức tạp đến mức khó có thể hiểu được người thám mã.

• **Độ khuếch tán – diffusion**. Tính chất có các mẫu thống kê trong bản rõ được phát tán khắp không gian bản mã.

**Thuật ngữ và ký hiệu cơ bản**

• P – bản rõ (*plaintext*). Là thông điệp có thể đọc được (được viết bằng ngôn ngữ tự nhiên).

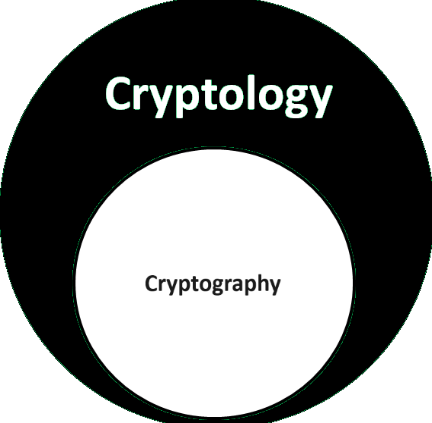
• C – bản mã (*ciphertext*). Là đầu ra của một lược đồ mã hóa và không thể đọc được.

• E – hàm mã hóa (*encryption*). Ví dụ, viết C = E(P), nghĩa là áp dụng hàm mã hóa E trên bản rõ P tạo ra bản mã C.

• D – hàm giải mã (*Decryption*), tức là D(C) = P. Lưu ý D[E(P)] = P và E[D(C)] = C.

**Mậ mã và mã hóa**

Như khái niệm nêu, mã hóa (*cryptography*) là “viết” dạng mã để bảo mật thông điệp cần trao đổi riêng tư, trong khi đó, mật mã (*cryptology*) nghiên cứu cả cách viết mã lẫn cách “đọc” mã. Hình dưới mô ta tương quan giữa hai khái niệm mã hóa và mật mã này.



Nói theo ngôn ngữ toán học và thuật toán thì mã hóa là thiết kế các thuật giải để dựng mã, còn mật mã là các thuật giải không chỉ dựng mã mà còn thám mã (*cryptanalysis*), nên ta có thể đúc kết bằng công thức tượng trưng

Cryptology = Cryptography + Crypanalysis ≡ Mật mã = Dựng Mã và Thám Mã.

• Mã hóa bao gồm các ứng dụng như bảo mật tin nhắn, truyền tập tin và chữ ký số.

• Mật mã bao gồm các ứng dụng như thám mã, phá mã và phân tích toán học các phương pháp mã hóa.

**HỆ MÃ**

**Mã hóa như một ánh xạ**

Ta có thể định nghĩa một hệ mã như một hàm “khả nghịch” E đi từ không gian các thông điệp hay bản rõ P vào không gian các bản mã C phụ thuộc vào một tham số trong không gian khóa K.

và ,

và một hàm D đi từ không gian các bản mã C vào lại không gian các bản rõ P

,

sao cho

.

. Hàm E, viết tắt của *Encryption*, được gọi là hàm mã hóa. Hàm E này phụ thuộc tham số k trong không gián khóa K (*key space*).

. k gọi là khóa mã hóa (*encryption key*).

. m được gọi là thông điệp (*message*) hay bản rõ (*plaintext*) thuộc không gian hữu hạn các bản rõ P (*message space, plaintext space*).

. c được gọi là bản mã (*ciphertext*) thuộc không gian các bản mã C (*ciphertext space*).

. Hàm D, viết tắt của *Decryption*, hàm giả mã. Hàm D phụ thuộc tham số k’ trong không gian khóa K.

. k’ gọi là khóa giải mã (*decryption key*).

**Mã hóa – một trò chơi**

Trong các ví dụ tiếp theo, ta đồng nhất mỗi ký tự trong tập chữ cái bằng thứ tự trong bảng chữ cái, nghĩa là A bằng 1, B là 2,…, Z là 26, và khoảng tráng ta cho bằng 0.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | | | | | | | | | | 1 | | | | | | | | | | 2 | | | | |
|  | A | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | X | Y | Z |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 |

Thi thông điệp gốc

CRYPTOGRAPHY IS A GAME

được ‘số hóa’ với mỗi ký tự bằng 2 số, thành

021723151914061701150723**00**0818**00**06011204

***Ví dụ 1***

,

Thì, với k = 3, thông điệp gốc mã hóa thành

052001182217092004181001**00**1121**00**09041507 ≡ FUASXRJUIESKA LV JEPH.

Sẽ được giải mã hoàn toàn với hàm:

.

Trò chơi này quá đơn giản, chỉ bằng ‘vét cạn’ thử qua 25 giá trị của khóa k trong tập K = {0, 1,…, 24}, ta có thể đọc được bản mã, là 1 trong 24 chuỗi giải được có ý nghĩa. Vì thế, luật chơi ở đây là giấu cả cách mã (*algorithm*) lẫn khóa (*key*).

**Luật chơi 1**: Giấu thuật E giải lẫn khóa k.

***Ví dụ 2***

Bây giờ, mỗi thông điệp m là một chữ số, , và hàm mã E với hàm giải mã D được định nghĩa bởi

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | m | D(c) |
| 6 | 7 | 1 | 3 | 9 | 8 | 0 | 2 | 4 | 5 | E(m) | c |

Khi ấy,

021723151914061701150723**00**0818**00**06011204 ↔ 61721378757960767786236664746660677169.

Trong trường hợp này k = 6713989245.

Nếu dùng cách ‘vét cạn’ như đã làm trong ví dụ 1, phải thử 10! giá trị khác nhau của k do có 10 cách để hoán vị 10 vị trí. Trong trường hợp tổng quá, ta có thể sử dụng khóa dài n vị trí, khi đó sẽ có n! giá trị khóa khác nhau, và sẽ không thể vét cạn nếu n lớn. Trong trường hợp này, luật chơi của ta có thể chỉ cần giữ bí mật khóa.

**Luật chơi 2**: công bố thuật giải E và giữ bí mật khóa k.

***Ví dụ 3***

Tương tự cách mã và giải mã như Ví dụ 3, chỉ khác là với

k = 0185947263.

Như vậy, có thể chơi theo luật chơi 2.

Nhận xét là khóa k ở đây có tính chất đặc biệt:

,

và ta có thể tính ngược lại

.

Như vậy ta có thể định nghĩa các hàm mã hóa và giải mã như sau:

,

và

.

Bây giờ ta có thể chơi theo luật mới mã hóa bằng một khóa (k = 3) và giải mã bằng khóa khác (k’=7).

**Luật chơi 3**: công bố thuật giải E, D và một phần khóa (k), giữ bí mật khóa giải mã (k’).

***Ví dụ 4***

Giả sử dùng hàm . Và hàm mã được định nghĩa đệ quy như sau:

.

Rõ ràng E không khả nghịch, và ta cũng có thể thấy (nếu chia dư cho một số nguyên tố p lớn, thay vì mod 11 như trong hàm σ), thì cho c = E(M), rất khó có thể tìm lại được M. Trò chơi này chỉ đơn giản là dùng hàm E được công bố để giấu dữ liệu M.

**Luật chơi 4**: Công bố thuật giải và không dùng khóa.

***Ví dụ 5***

Tương tự các ví dụ 2 và 3, chỉ khác hàm và khóa là hoán vị 0123456789 ↔ 01643972875. Hàm hoán vị này có tính chất:

.

Tính chất đặc biệt này có thể xem như nguyên lý tạo hệ mã.

**Các loại hệ mã**

Sử dụng tính chất khóa cũng như tính chất tồn tại hay không hàm giải mã để phân loại các hệ mã như sau. Trước hết, theo tính chất khóa.

***Mã đối xứng***

Mã đối xứng (*symmetric cryptosystem*), là hệ mã tương ứng với trường hợp khóa k tương đương với khóa k’, k ≡ k’, nghĩa là cho k có thể suy ra k’ và ngược lại.

Rõ ràng khóa k phải được giữ bí mật giữa các thành phần liên quan đến thông điệp mật, nên hệ mã dạng này còn được gọi là mã khóa bí mật, hay mã bí mật (*secret key cryptosystem*).

Trong thực tế, để chỉ 2 đối tác hiểu được tin mật trao đổi giữa họ, khóa k phải được quy ước hay chia sẻ trước (*pre-share key*).

Ví dụ 2 ở trên, dùng luật chơi 2, là minh họa cho dạng hệ mã đối xứng này. Ở đó, khóa bí mật k chính là chuỗi hoán vị của chuỗi n = 10 kí hiệu, đánh nhãn từ 0 đến 9.

***Mã bất đối xứng***

Mã bất đối xứng (asymmetric cryptosystem), là hệ mã tương ứng với trường hợp k ≠ k’. Nghĩa là mã hóa bằng một khóa k và giải mã bằng một khóa k’ khác k. Trong trường hợp này, khóa k được ký hiệu là e, và k’ được ký hiệu d.

Do mã và giải mã không cùng khóa nên khóa mã hóa có thể được công bố công khai để ai cũng có thể gửi tin mật cho người có khóa d, và chỉ người giữ khóa d mới có thể giải mã được bản mã mã hóa dùng khóa e. Khóa d phải được giữ bí mật. Ví thế, hệ mã trong trường hợp này còn được gọi là hệ mã khóa công khai, hay gọi đơn giản là hệ mã công khai (*public key cryptosystem*), và e được gọi là khóa công khai (*public key*) còn d là khóa cá nhân (*private key*).

Ví dụ 3 ở trên, dùng luật chơi 3, là một minh họa đơn giản của hệ mã công khai, ở đó, khóa e được công bố công khai, và khóa cá nhân d phải được giữ bí mật.

Bây giờ, không quan tâm đến tính chất ‘khả nghịch’của hàm mã, nghĩa là ta chỉ cần giấu chứ không có nhu cầu giải mã.

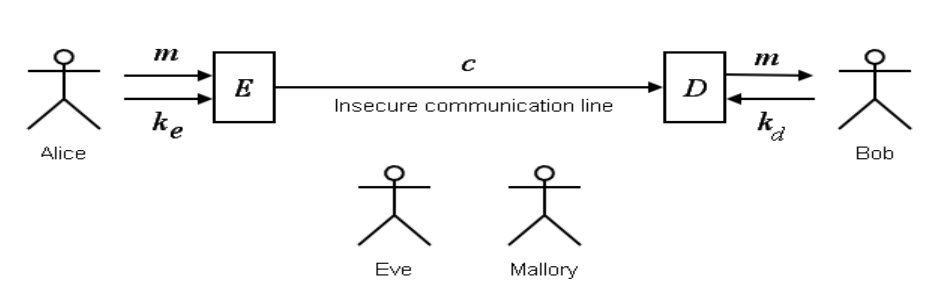
***Hàm băm mật mã***

Hàm băm mật mã (*cryptographic hash function*) là hệ mã ở đó ta chỉ có nhu cầu bảo mật tập các thông điệp trong cùng 1 bản mã chứ không có nhu cầu giải mã. Thường thì bản mã có chiều dài theo số bit ngắn hơn rất nhiều so với tập nhiều thông điệp, vì thế hàm được gọi là hàm băm (*hash function*), nghĩa là nó biểu diễn một văn bản, là tập các thông điệp, bằng một mã hay giá trị băm (*hash value*).

Ví dụ 4 ở trên minh họa cho hệ mã này.

**Các tính chất của hệ mã**

Hình sau minh họa hoạt động của một hệ thống trao đổi thông tin bảo mật trên các kênh truyên không an toàn (*unsecure channel*), trong đó, Alice và Bob là 2 đối tác có nhu cầu trao đổi tin mật với nhau trên kênh truyền không an toàn có cả Eve, Mallory, nhưng ngoài Alice và Bob, các người khác, bao gồm cả Eve và Mallory, không thể hiểu được trao đổi giưa Alice và Bob.



Một hệ mật an toàn phải có những tính chất sau:

*Tính bảo mật* – **confidentiality**. Chi phí để Eve khôi phục lại thông điệp m từ bản mã c sẽ cao hơn giá trị của thông điệp m.

*Xác thực –* **authentication**. Bob có thể dễ dàng xác minh là chính Alice đã gửi thông điệp m.

*Tính toàn vẹn* – **integrity**. Bob có thể xác minh rằng thông điệp m không bị giả mạo hay thay đổi gì.

*Không từ chối –* **non-repudiation**. Alice không thể từ chối là mình không gửi thông điệp m.

Các hệ mã tốt nhất giả định rằng Eve và Mallory ngay cả biết hàm mã E, hàm giải mã D, bản mã c và biết khóa công khai nếu hệ là mã công khai. Hầu hết các hệ thống mật mã không dựa vào tính bí mật của thuật toán, vì những lý do sau:

* Nếu thuật toán bí mật của bạn bị xâm phạm, sẽ phải thay thuật toán mới, khó hơn nhiều việc chỉ thay một khóa!
* Công khai thuật toán có thể nhận được sự hỗ trợ của hàng nghìn chuyên gia tìm kiếm lỗi.

***Shannon's Maxim***

Nguyên lý Maxim Shannon giả thiết rằng kẻ thù biết hệ thống! Thuật toán bảo mật phải bền vững trong ngữ cảnh kẻ thù biết mọi thứ về hệ thống ngoại trừ khóa.

Nghiên cứu về mã hóa (cryptography) học bao gồm thiết kế các loại mật mã khác nhau, trao đổi khóa, khóa xác thực, hàm băm mật mã, chữ ký số và các vấn đề xã hội (pháp lý, chính trị, v.v.). Tuy vậy, ta cũng sẽ học một số phương pháp phân thám mã (tấn công) để đảm bảo hệ mã xây dựng hay đang sử dụng là an toàn-bảo mật.

**Các phương pháp mã hóa**

Dưới đây là một số phương pháp mã hóa. Lưu ý rằng một hệ mã thường có nhiều hơn một trong các phương pháp sau.

***Phương pháp***

*Thay thế* – **substitution** . Mỗi ký hiệu của bản rõ được thay thế bằng một hoặc nhiều ký hiểu để tạo ra bản mã

*Chuyển vị* – **transposition**. Các ký tự trong bản rõ được sắp xếp lại để tạo thành bản mã

*Thay thế đơn điệu* – **Monoalphabetic**. Mã thay thế trong đó một ký hiệu của bản rõ luôn được thay thế bằng cùng một ký hiệu.

*Thay thế đa chiều* – **polyalphabetic**. Mã thay thế về cơ bản sử dụng nhiều ánh xạ thay thế dùng một bảng chữ cái.

*Tuần hoàn* – **periodic**. Mã thay thế đa chiều trong đó sơ đồ thay thế được lặp lại.

*Không-tuần hoàn* – **non-periodic.**

*Mã khối* – **block cipher**. Mã hóa diễn theo từng khối ký hiệu.

*Mã Luồng* – **stream cipher**. Mã hoạt động trên luồng dữ liệu có độ dài không xác định, thường kết hợp phản hồi.

LƯU Ý: Trong các ví dụ dựa trên ký tự bên dưới, chúng tôi sẽ giả sử (không mất tính tổng quát) bảng chữ cái 26 ký hiệu (A..Z).

***Ví dụ***

*Hệ mã thay thế đơn điệu* – **Monoalphabetic**

Thay vì chỉ thêm một khoảng bù cố định cho mỗi ký tự, có thể tính trước bảng thay thế bằng cách tạo một hoán vị ngẫu nhiên cho bảng chữ cái của mình. Chẳng hạn:

ABCDEFGHIJKLMNOPQRSTUVWXYZ

MQHPSVJYCURFTBILAKWNGZDOEX

Thì thông điệp ATTACKATDAWN được mã hóa thành MNNMHRMNPMDB.

Không thể giải mã bằng cách đoán khóa (có n! khóa có thể), nhưng phân tích tần số có thể giải mã bất kỳ khóa nào của mã thay thế đơn điêu, miễn là thông điệp đủ dài.

Đối với các kỹ thuật có khóa là hoán vị, một cách để làm cho khóa dễ nhớ hơn là chọn một cụm từ, sắp xếp nó các chữ cái duy nhất, sau đó điền các chữ cái còn thiếu theo thứ tự. Chẳng hạn, từ

PREMATURE OPTIMIZATION IS THE ROOT OF ALL EVIL

tạo ra ánh xạ thay thế:

PREMATUOIZNSHFLVBCDGJKQWXY

**Các phương pháp mã hóa hiện đại**

Các kỹ thuật hiện đại gồm

• Thi hành trên chuỗi bit, thay vì trên chuỗi ký tự như các ví dụ ở trên.

• Che giấu hoàn toàn các mẫu và phần dư thừa trong bản rõ.

• Sử dụng khóa ngẫu nhiên (cũng có thể được sử dụng lại).

• Đảm bảo rằng những thay đổi rất nhỏ trong bản rõ sẽ ảnh hưởng đến phần lớn bản mã (và ngược lại). Đây được gọi là hiệu ứng Avalanche. Ngoài ra, hệ mã có khả năng chịu lỗi hiệu quả.

**TẤN CÔNG HỆ MÃ**

Tấn công hệ mã nhằm mục đích hiểu được nhừng già đã được mã hóa, hay chí ít, lấy được một số thông tin hữu ích của hệ mã. Có thể tấn công trực tiếp vào thuật toán của hệ mã bằng cách phân tích các thuật toán của hệ mã, gọi là thám mã (*cryptanalysis*), hay tấn công vào giao thức bằng công nghệ, gọi là phá mã (*hack*).

**Thám mã**

***Tấn công biết bản rõ***

Tấn công biết bản rõ – **KPA** (*known plaintext attack*). Là mô hình thám mã mà kẻ tấn công có quyền truy cập vào cả văn bản gốc lẫn bản mã. KPA được sử dụng khi muốn lấy thêm thông tin bí mật như khóa bí mật và bảng từ mã.

***Tấn công biết bản mã***

Tấn công biết bản mã – **COA** (*Ciphertext Only Attack* or *known ciphertext attack*). là mô hình thám mã mà kẻ tấn công chỉ có quyền truy cập vào một tập các bản mã. Mặc dù kẻ tấn công không có truy cập vào bản rõ trước khi mã hóa, nhưng trong COA, kẻ tấn công vẫn có một số kiến thức về bản rõ. Chẳng hạn, kẻ tấn công có thể biết ngôn ngữ gốc của bản rõ là tiếng Việt hay tiếng Anh, hoặc biết phân phối của các ký tự trong bản rõ. Dữ liệu và thông điệp của giao thức chuẩn thường là một phần của bản rõ trong nhiều hệ thống và thường có thể được đoán hoặc biết như một phần của tấn công biết bản mã vào các hệ thống này.

***Tấn công chọn bản rõ***

Tấn công chọn bản rõ – CPA (*chosen plaintext attack*). Là mô hình thám mã, giả định rằng kẻ tấn công có thể lấy được bản mã cho các bản rõ tùy ý. Mục tiêu của tấn công là lấy thông tin làm giảm tính bảo mật của sơ đồ mã hóa.

Các hệ mã hiện đại có tính chất bảo mật ngữ nghĩa, còn được gọi là tính không thể phân biệt được của mã khi thực hiên CPA và do đó, theo thiết kế, chúng thường kháng được các CPAs.

***Tấn công chọn khóa***

Tấn công chọn khóa – **CKA** (*chosen key attack*). Là mô hình thám mã trong đó người tấn công có thể thu thập thông tin bằng cách lấy được bản giải mã của các bản mã đã được chọn. Từ những thông tin này, kẻ tấn công tìm cách khôi phục khóa bí mật ẩn được sử dụng để giải mã.

***Thám mã tuyến tính***

Thám mã tuyến tính – **LC** (*linear cryptanalysis*). Là một hình thức thám mã tổng quát dựa trên xấp xỉ affine hoạt động của hệ mã. Thám mã tuyến tính là một trong hai phương pháp tấn công được sử dụng rộng rãi nhất đối với mã khối (*block cipher*) còn gọi lại là thám mã mã sai hân.

***Thám mã vi sai***

Thám mã vi sai – **DC** (*diferential cryptanalysis*). Là một dạng của thám mã tổng quát được áp dụng chủ yếu tấn công mã khối, mã dòng và hàm băm mật mã. Ý tưởng là nghiên cứu những sự khác biệt trong thông tin đầu vào ảnh hưởng thế nào đến sự đầu ra. Trong trường hợp mã khối, đó là các kỹ thuật để tìm sự khác biệt thông qua mạng biến đổi, khám phá nơi mật mã thể hiện hành vi không ngẫu nhiên và khai thác các thuộc tính đó để khôi phục khóa bí mật (khóa mật mã).

**Đánh cắp thông tin**

Đánh cắp thông tin – data theft. Bao gồm thông tin mật khẩu, mã phần mềm hoặc thuật toán cũng như các quy trình hoặc công nghệ độc quyền. Trộm cắp dữ liệu được coi là hành vi vi phạm về bảo mật và quyền riêng tư, có khả năng gây hậu quả nghiêm trọng cho các cá nhân và tổ chức.

trộm cắp dữ liệu là hành động đánh cắp thông tin kỹ thuật số được lưu trữ trên máy tính, máy chủ hoặc thiết bị điện tử để lấy thông tin bí mật hoặc xâm phạm quyền riêng tư. Dữ liệu bị đánh cắp có thể là bất cứ thứ gì từ thông tin tài khoản ngân hàng, mật khẩu trực tuyến, số hộ chiếu, số giấy phép lái xe, số an sinh xã hội, hồ sơ y tế, đăng ký trực tuyến, v.v. Khi một người không được ủy quyền có quyền truy cập vào thông tin cá nhân hoặc tài chính, họ có thể xóa, thay đổi hoặc ngăn chặn quyền truy cập vào thông tin đó mà không có sự cho phép của chủ sở hữu.

Các thuật ngữ ‘xâm phạm dữ liệu’ và 'rò rỉ dữ liệu' có thể được sử dụng như nhau khi thảo luận về hành vi đánh cắp dữ liệu. Tuy nhiên, chúng khác nhau:

* Rò rỉ dữ liệu xảy ra khi dữ liệu nhạy cảm vô tình bị lộ trên internet hoặc thông qua ổ cứng hoặc thiết bị bị mất. Qua đó tin tặc có được quyền truy cập trái phép vào dữ liệu nhạy cảm.
* Ngược lại, xâm phạm dữ liệu đề cập đến các cuộc tấn công có chủ ý.

Các phương pháp tấn công đánh cắp gồm:

***Kỹ thuật xã hội***

Kỹ thuật xã hội (*social engineering*) phổ biến nhất là lừa đảo. Lừa đảo xảy ra khi kẻ tấn công giả dạng một thực thể đáng tin cậy để lừa nạn nhân mở email, tin nhắn văn bản hoặc tin nhắn tức thời. Bị lừa đảo dẫn đến việc đánh cắp dữ liệu.

***Mật khẩu yếu***

Mật khẩu yếu (*weak password*). Là sử dụng mật khẩu dễ đoán hoặc sử dụng cùng một mật khẩu cho nhiều tài khoản. Thói quen sử dụng mật khẩu kém – chẳng hạn như viết mật khẩu ra một tờ giấy hoặc chia sẻ chúng với người khác – cũng có thể dẫn đến việc đánh cắp dữ liệu.

***Lỗ hổng hệ thống***

Lỗi hệ thống (*system errors*). Là do ứng dụng hoặc hệ thống mạng được viết kém, được thiết kế hoặc triển khai kém sẽ tạo ra các lỗ hổng mà tin tặc có thể khai thác và sử dụng để đánh cắp dữ liệu. Phần mềm chống vi-rút đã lỗi thời cũng có thể tạo ra lỗ hổng.

***Rủi ro từ bên trong***

Nhân viên làm việc cho một tổ chức có quyền truy cập vào thông tin cá nhân của khách hàng có thể sao chép, thay đổi hoặc đánh cắp dữ liệu. Các rủi ro này không chỉ giới hạn ở các nhân viên hiện tại. Họ cũng có thể là cựu nhân viên hoặc đối tác có quyền truy cập vào hệ thống hoặc thông tin nhạy cảm của tổ chức.

***Lỗi của con người***

Xâm phạm dữ liệu không chỉ là kết quả của các hành động độc hại. Đôi khi là lỗi của con người. Như gửi thông tin nhạy cảm đến nhầm người, đính kèm sai tài liệu hoặc giao tập tin cho người không có quyền truy cập. Ngoài ra, lỗi của con người có thể liên quan đến việc cấu hình sai, chẳng hạn như nhân viên để lại cơ sở dữ liệu chứa thông tin nhạy cảm trực tuyến mà không có bất kỳ hạn chế nào về mật khẩu.

***Tải xuống bị xâm phạm***

Tải xuống các chương trình hoặc dữ liệu từ các trang web bị nhiễm vi-rút hoặc phần mềm độc hại. Điều này cho phép tin tặc truy cập trái phép vào thiết bị để đánh cắp dữ liệu.

***Hành động thể chất***

Một số hành vi đánh cắp dữ liệu là các hành động vật lý bao gồm trộm cắp giấy tờ hoặc thiết bị như máy tính xách tay, điện thoại hoặc thiết bị lưu trữ, hay xem lén màn hình lúc đăng nhập. Đánh cắp thẻ - nơi thiết bị được đưa vào đầu đọc thẻ và máy ATM để thu thập thông tin thẻ thanh toán - là một nguồn trộm dữ liệu khác.

***Sự cố về cơ sở dữ liệu hoặc máy chủ***

Nơi lưu trữ thông tin bị tấn công do sự cố về cơ sở dữ liệu hoặc máy chủ, kẻ tấn công có thể truy cập thông tin cá nhân của khách hàng.

***Thông tin có sẵn công khai***

Rất nhiều thông tin có thể được tìm thấy trong mạng xã hội - thông qua tìm kiếm trên internet và xem qua các bài đăng của người dùng trên mạng xã hội, có thể được khai thác cho mục đích tấn công.