2 - Cifrado de flujo ChaCha20

April 2, 2023

1 Tema 3 - Cifrado de flujo con ChaCha20

En Python tenemos dos posibilidades para probar criptografía: el paquete `PyCryptodome` y el paquete `cryptography`. Ambos son opciones válidas. Las prácticas de este curso las haremos con PyCryptodome. Puedes encontrar la ayuda en: https://pycryptodome.readthedocs.io/en/latest/

Empezamos instalando PyCryptotome e importando lo que vamos a necesitar:

[146]:
```
!python3 -m pip install pycryptodome
from base64 import b64encode, b64decode
from Crypto.Cipher import ChaCha20
from Crypto.Random import get_random_bytes
```

```
Requirement already satisfied: pycryptodome in
/home/gattes/.cache/pypoetry/virtualenvs/jupyter-notebooks-QRvsWska-
py3.10/lib/python3.10/site-packages (3.17)

[notice] A new release of pip is
available: 23.0 -> 23.0.1
[notice] To update, run:
pip install --upgrade pip
```

1.1 Cifrado y envío de datos

Los módulos de criptografía suelen necesitar una etapa inicial de configuración. Cada módulo se configura a su manera. A continuación encontrarás la etapa de configuración de ChaCha20 para PyCryptodome.

Fíjate que la clave se crea al azar con algoritmos criptográficos `Crypto.Random.get_random_bytes()`: **es fundamental que las claves sean totalmente aleatorias y creadas también con algoritmos criptográficos**. Como habrás visto en el ejercicio "creando azar" de este mismo tema, no todas las funciones de creación de azar son válidas.

[147]:
```
key = get_random_bytes(32)
cipher_emisor = ChaCha20.new(key=key, nonce=None)
print('Longitud de la clave: {} bits'.format(8 * len(key)))
```

```
Longitud de la clave: 256 bits
```

En PyCryptodome el *nonce* se puede pasar al algoritmo durante la configuración. Si, como en este caso, no se pasa *nonce* durante la creación, la librería crea un *nonc* al azar que podemos recuperar. Si decides crear tú el *nonce*, recuerda que también tiene que ser un número aleatorio creado con algoritmos criptográficos, igual que la clave.

[148]:
```
nonce = b64encode(cipher_emisor.nonce)
print('Longitud del nonce: {} bits'.format(8 * len(cipher_emisor.nonce)))
print(nonce)
```

```
Longitud del nonce: 64 bits
b'ZJnr/upa4to='
```

El emisor cifra el mensaje `Atacaremos al amanecer` y envía al receptor `result`, es decir, tanto como mensaje cifrado como el *nonce*. Fíjate: el nonce se puede enviar por un canal inseguro, así que se asume que el atacante lo conocerá.

Observa que el resultado lo codificamos en Base64 (https://es.wikipedia.org/wiki/Base64). Aunque no es necesario, sí que es común hacerlo así porque algunos protocolos (correo electrónico, JSON…) solo puede enviar caracteres imprimibles. No pierdes ni ganas seguridad si decides usar o no Base64, es más una exigencia de tu sistema de comunicaciones. Fíjate que he usado la expresión "codificamos en Base64", no ciframos. Base64 es un algoritmo de codificación de bytes, no tiene claves, cualquier lo puede codificar y decodificar y por tanto no es un cifrado.

[149]:
```
plaintext = b'Atacaremos al amanecer'
ciphertext = cipher_emisor.encrypt(plaintext)
ct = b64encode(ciphertext)
result = {'nonce':nonce, 'ciphertext':ct}
print(result)
```

```
{'nonce': b'ZJnr/upa4to=', 'ciphertext': b'V+ypuNeAtpilaVtAz9CJfFry+Uuaaw=='}
```

1.2 Recepción y descifrado

El receptor toma el *nonce* y el *ciphertext*, primero decodifica el base64, configura el *cipher* con la clave que conoce (ya veremos cómo la conoce en el tema 4 y 5) y el *nonce* que ha recibido y descifra:

[150]:
```
received_nonce = b64decode(result['nonce'])
received_ciphertext = b64decode(result['ciphertext'])
cipher_receptor = ChaCha20.new(key=key, nonce=received_nonce)
plaintext = cipher_receptor.decrypt(received_ciphertext)
print(plaintext)
```

```
b'Atacaremos al amanecer'
```

1.3 Siguientes mensajes: sincronización entre ciphers

Supongamos que el usuario vuelve a enviar el mismo mensaje, con el mismo cipher (fíjate que no volvemos a definir `cipher_emisor`: lo estamos reaprovechando)

[151]:
```
plaintext = b'Atacaremos al amanecer'
ciphertext = cipher_emisor.encrypt(plaintext)
ct = b64encode(ciphertext)
result = {'nonce':nonce, 'ciphertext':ct}
print(result)
```

```
{'nonce': b'ZJnr/upa4to=', 'ciphertext': b'D7utmvv7jol6epl2d4b1OCpZOOzRMg=='}
```

Fíjate: estamos cifrando el mismo mensaje con el mismo nonce… pero el ciphertext es diferente. ¿Recuerdas que nunca se debe cifrar el mismo texto con la misma clave? ChaCha20 nos ayuda a que no lo hagamos, ni siquiera por equivocación, mediante el uso de un contador.

Supongamos que el receptor crea un nuevo cipher, con la misma configuración de key y nonce, e intenta descifrar:

[152]:
```
received_nonce = b64decode(result['nonce'])
received_ciphertext = b64decode(result['ciphertext'])
cipher_receptor = ChaCha20.new(key=key, nonce=received_nonce)
plaintext = cipher_receptor.decrypt(received_ciphertext)
print(plaintext)
```

```
b'\x19#eAM\t]|\xb0`\xe2W\xd4v\x1d)\x11\xc5L\xc4.+'
```

¿Qué ha pasado? ¿Por qué no se descifra? Recuerda que ChaCha20 tiene un contador adicional interno. Es decir: **emisor y receptor tienen que estar sincronizados** Es decir: para descifrar el byte número 22 tenemos que decirle al receptor que han pasado 22 bytes antes, aunque no los haya visto.

(nota: 22 es el tamaño en bytes de la cadena “Atacaremos al amanecer”, que fue el contenido del primer mensaje)

Si volvemos a intentar descifrar, ahora sí que podemos hacerlo:

[153]:
```
cipher_receptor.seek(22)
plaintext = cipher_receptor.decrypt(received_ciphertext)
print(plaintext)
```

```
b'Atacaremos al amanecer'
```

PyCryptodome y todos los demás están sincronizados siempre que descifremos los mismos bytes que hemos cifrado desde que se han creado los dos ciphers, el de emisión y el de recepción.

Si alguno de los dos pierde la sincronización (por ejemplo, porque se reinicia), entonces es necesario volver a sincronizarlos con un “seek”: “ya envié XX bytes aunque no los hayas visto, mueve el estado a esta posición”

Poder volver a sincronizar los dos streams es una enorme ventaja de ChaCha20 y eso es por el parámetro `pos` autoincremental que forma parte de la matriz de estado. No todos los algoritmos permiten sincronizar los flujos si se pierde la sincronización.