How crypto fails in practice? CSS and WEP

*Slides borrowed from Vitaly Shmatikov

Stream Ciphers

- One-time pad:

  Ciphertext(Key,Message)=Message⊕Key

  - Key must be a random bit sequence as long as message
- Idea: replace "random" with "pseudo-random"
  - Use a pseudo-random number generator (PRNG)
  - PRNG takes a short, truly random secret seed and expands it into a long "random-looking" sequence
    - E.g., 128-bit seed into a 10^6-bit pseudo-random sequence

No efficient algorithm can tell this sequence from truly random

- Ciphertext(Key,Msg)=IV, Msg⊕PRNG(IV,Key)
  - Message processed bit by bit (unlike block cipher)

Stream Cipher Terminology

- The seed of a pseudo-random generator typically consists of initialization vector (IV) and key
  - The key is a secret known only to the sender and the recipient, not sent with the ciphertext
  - IV is usually sent with the ciphertext
- The pseudo-random bit stream produced by PRNG(IV,key) is referred to as the keystream
- Encrypt message by XORing with keystream
  - ciphertext = message $\oplus$ keystream

Properties of Stream Ciphers

- Usually very fast (faster than block ciphers)
  - Used where speed is important: WiFi, DVD, RFID, VoIP
- Unlike one-time pad, stream ciphers do not provide perfect secrecy
  - Only as secure as the underlying PRNG
  - If used properly, can be as secure as block ciphers
- PRNG must be cryptographically secure

Using Stream Ciphers

- No integrity
  - Associativity & commutativity:
    $(M_1 \oplus PRNG(seed)) \oplus M_2 = (M_1 \oplus M_2) \oplus PRNG(seed)$
  - Need an additional integrity protection mechanism
- Known-plaintext attack is very dangerous if keystream is ever repeated
  - Self-cancellation property of XOR: $X \oplus X = 0$
  - $(M_1 \oplus PRNG(seed)) \oplus (M_2 \oplus PRNG(seed)) = M_1 \oplus M_2$
  - If attacker knows M_1, then easily recovers M_2 … also, most plaintexts contain enough redundancy that can recover parts of both messages from $M_1 \oplus M_2$

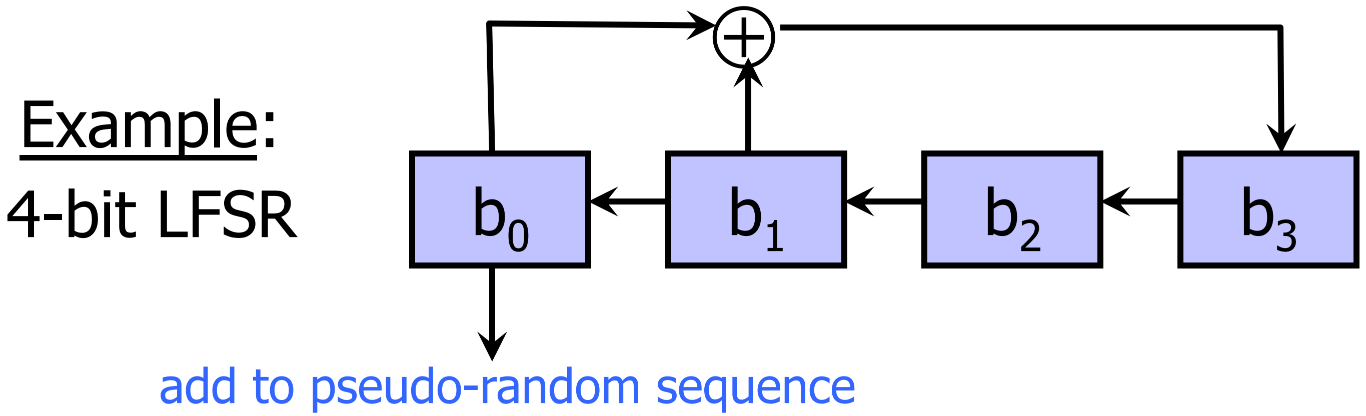
How Random is “Random”?

Cryptographically Secure PRNG

- Next-bit test: given N bits of the pseudo-random sequence, predict $(N+1)^{st}$ bit
  - Probability of correct prediction should be very close to 1/2 for any efficient adversarial algorithm

    (means what?)
- PRNG state compromise
  - Even if the attacker learns the complete or partial state of the PRNG, he should not be able to reproduce the previously generated sequence
    - … or future sequence, if there'll be future random seed(s)
- Common PRNGs are <u>not</u> cryptographically secure

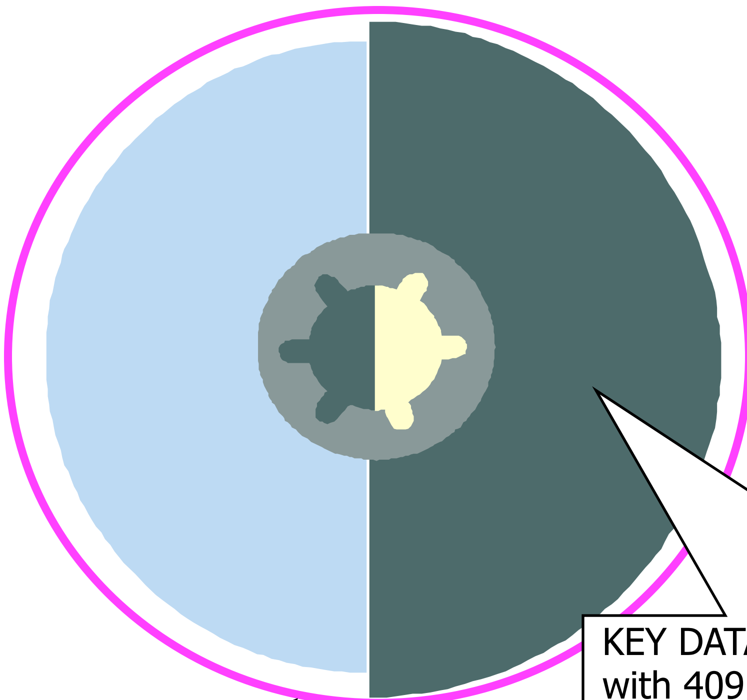
LFSR: Linear Feedback Shift Register

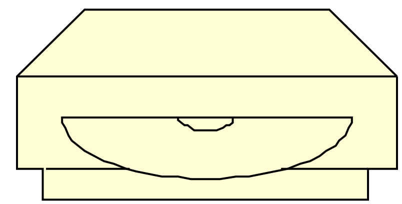


- For example, if the seed is 1001, the generated sequence is 1001101011110001001…
- Repeats after 15 bits (2^4-1)

Content Scrambling System (CSS)

- DVD encryption scheme from Matsushita and Toshiba





Each player has its own PLAYER KEY
(409 player manufacturers,
each has its player key)

KEY DATA BLOCK contains disk key encrypted
with 409 different player keys:

- $\text{Encrypt}_{\text{DiskKey}}(\text{DiskKey})$
- $\text{Encrypt}_{\text{PlayerKey1}}(\text{DiskKey})$ … $\text{Encrypt}_{\text{PlayerKey409}}(\text{DiskKey})$

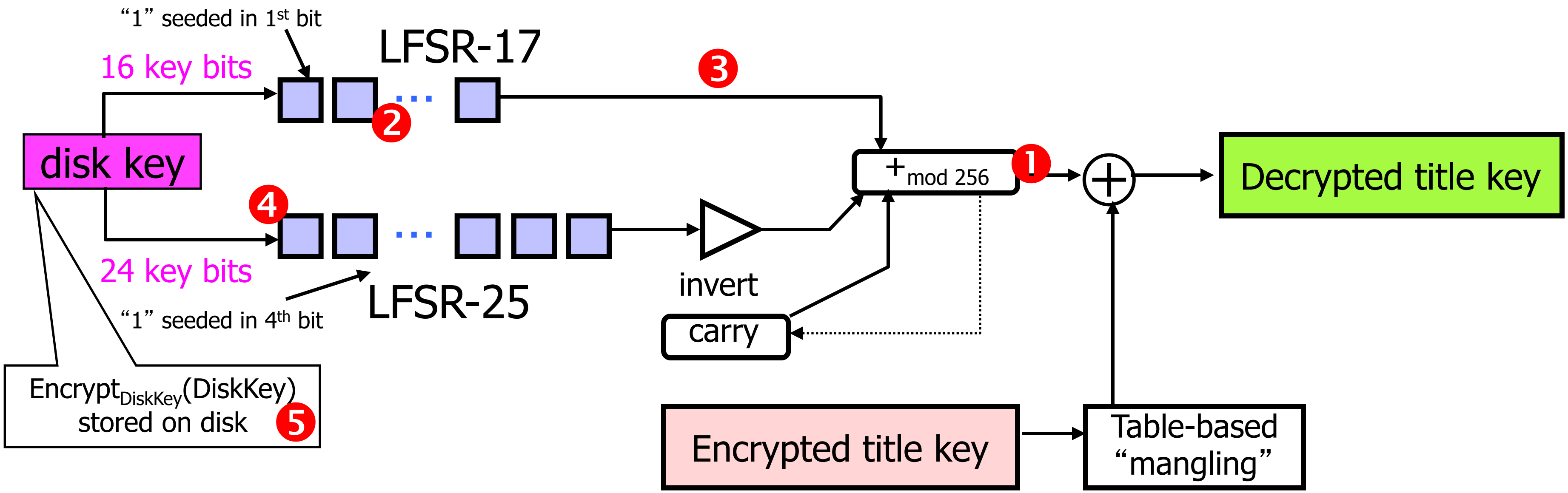
Each DVD is encrypted with
a disk-specific 40-bit DISK KEY

This helps attacker
verify his guess of disk key

What happens if even a single
player key is compromised?

Attack on CSS Decryption Scheme

[Frank Stevenson]



❶ Given known 40-bit plaintext, repeat the following 5 times (once for each plaintext byte): guess the byte output by the sum of the two LFSRs; use known ciphertext to verify – this takes $O(2^8)$

❷ For each guessed output byte, guess 16 bits contained in LFSR-17 – this takes $O(2^{16})$

❸ Clock out 24 bits out of LFSR-17, use subtraction to determine the corresponding output bits of LFSR-25 – this reveals all of LFSR-25 except the highest bit

❹ "Roll back" 24 bits, try both possibilities – this takes $O(2)$

❺ Clock out 16 more bits out of both LFSRs, verify the key

This attack takes $O(2^{25})$

DeCSS

- In CSS, disk key is encrypted under hundreds of different player keys… including Xing, a software DVD player
- Reverse engineering the object code of Xing revealed its player key
  - Every CSS disk contains the master disk key encrypted under Xing's key
  - One bad player $\Rightarrow$ entire system is broken!
- Easy-to-use DeCSS software

DeCSS Aftermath

- DVD CCA sued Jon Lech Johansen ("DVD Jon"), one of DeCSS authors - eventually dropped
- Publishing DeCSS code violates copyright
  - Underground distribution as haikus and T-shirts
  - "Court to address DeCSS T-Shirt: When can a T-shirt become a trade secret? When it tells you how to copy a DVD." - Wired News

RC4

- Designed by Ron Rivest for RSA in 1987
- Simple, fast, widely used
  - SSL/TLS for Web security, WEP for wireless

```
Byte array S[256] contains a permutation of numbers from 0 to 255
i = j := 0
loop
    i := (i+1) mod 256
    j := (j+S[i]) mod 256
    swap(S[i],S[j])
    output (S[i]+S[j]) mod 256
end loop
```

RC4 Initialization

```
Divide key K into L bytes
for i = 0 to 255 do
    S[i] := i
j := 0
for i = 0 to 255 do
        j := (j+S[i]+K[i mod L]) mod 256
        swap(S[i],S[j])
```

Key can be any length up to 2048 bits

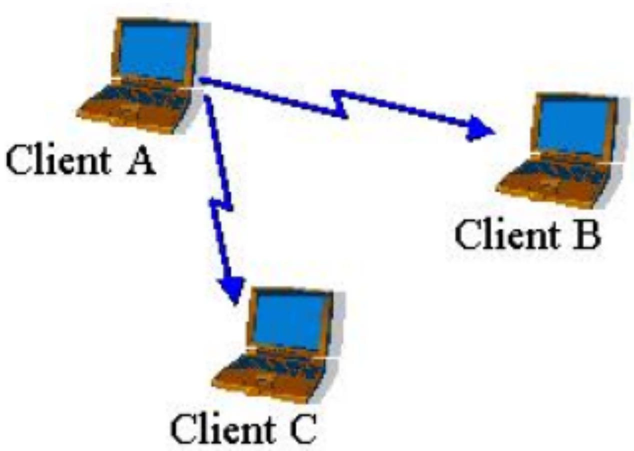
Generate initial permutation from key K

- To use RC4, usually prepend initialization vector (IV) to the key
  - IV can be random or a counter
- RC4 is not random enough… First byte of generated sequence depends only on 3 cells of state array S - this can be used to extract the key!
  - To use RC4 securely, RSA suggests discarding first 256 bytes

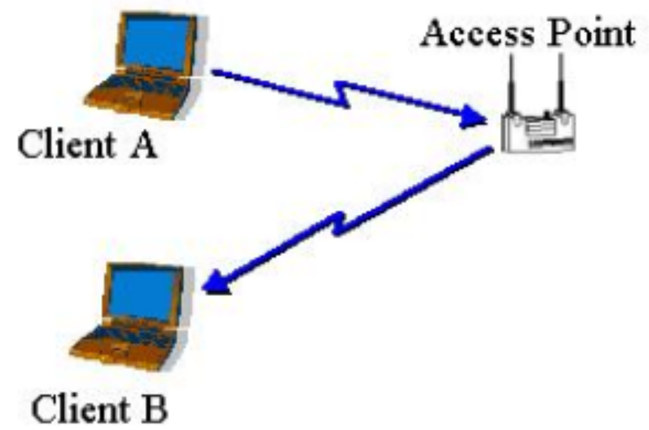
Fluhrer-Mantin-Shamir attack

802.11b Overview

- Standard for wireless networks (IEEE 1999)
- Two modes: infrastructure and ad hoc



IBSS (ad hoc) mode



BSS (infrastructure) mode

Access Point SSID

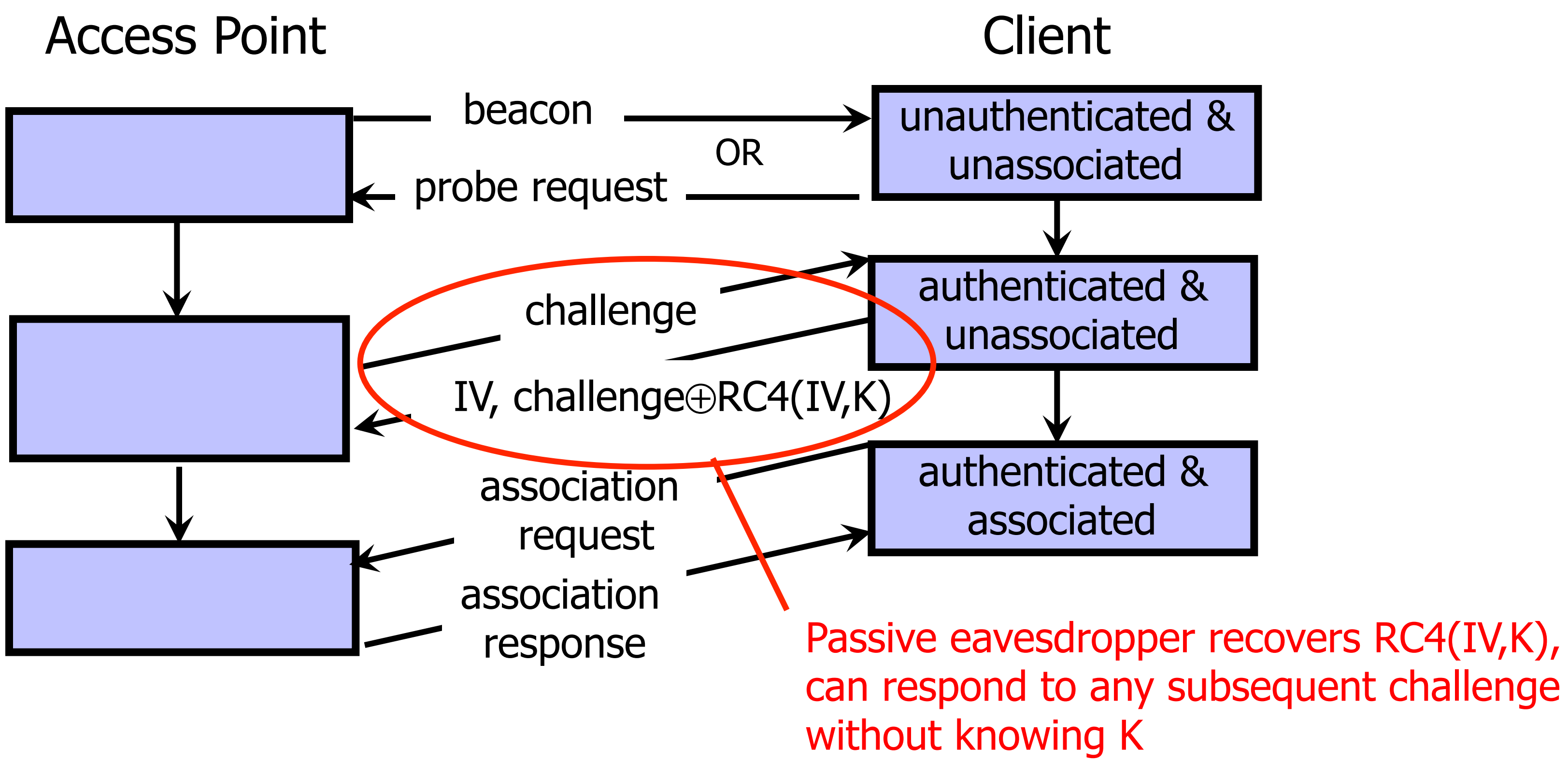
- Service Set Identifier (SSID) is the "name" of the access point
  - By default, access point broadcasts its SSID in plaintext "beacon frames" every few seconds
- Default SSIDs are easily guessable
  - Manufacturer's defaults: "linksys", "tsunami", etc.
  - This gives away the fact that access point is active
- Access point settings can be changed to prevent it from announcing its presence in beacon frames and from using an easily guessable SSID
  - But then every user must know SSID in advance

WEP: Wired Equivalent Privacy

- Special-purpose protocol for 802.11b
- Goals: confidentiality, integrity, authentication
  - Intended to make wireless as secure as wired network
- Assumes that a secret key is shared between access point and client
- Uses RC4 stream cipher seeded with 24-bit initialization vector and 40-bit key
  - Terrible design choice for wireless environment

Shared-Key Authentication

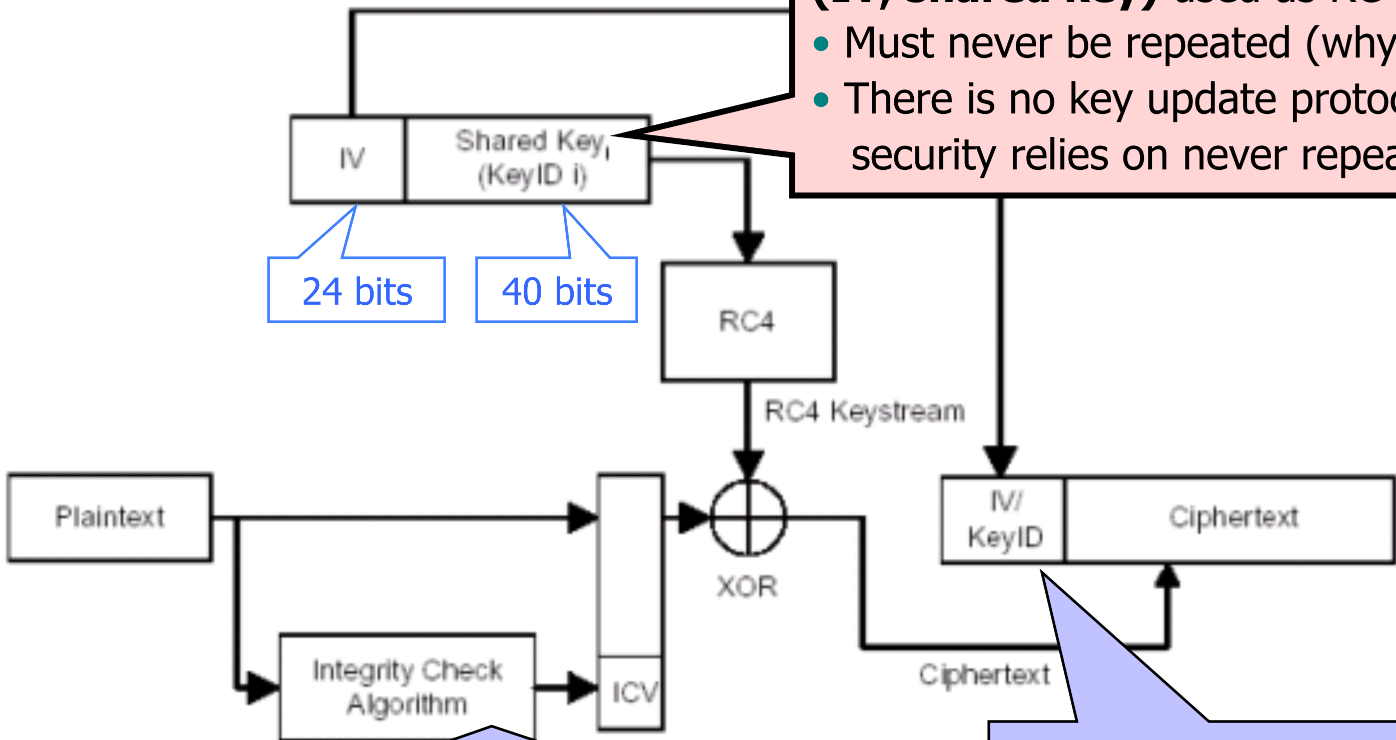
Prior to communicating data, access point may require client to authenticate



How WEP Works

(IV, shared key) used as RC4 seed

- Must never be repeated (why?)
- There is no key update protocol, so security relies on never repeating IV



IV sent in the clear

Worse: changing IV with each packet is optional!

CRC-32 checksum is linear in ⊕:
if attacker flips some plaintext bits, he knows which bits of CRC to flip to produce the same checksum

no integrity!

RC4 Is a Bad Choice for Wireless

- Stream ciphers require sender and receiver to be at the same place in the keystream
  - Not suitable when packet losses are common
- WEP solution: a separate keystream for each packet (requires a separate seed for each packet)
  - Can decrypt a packet even if a previous packet was lost
- But there aren't enough possible seeds!
  - RC4 seed = 24-bit initialization vector + <u>fixed</u> key
  - Assuming 1500-byte packets at 11 Mbps, 2^{24} possible IVs will be exhausted in about 5 hours
- Seed reuse is deadly for stream ciphers

Recovering the Keystream

- Get access point to encrypt a known plaintext
  - Send spam, access point will encrypt and forward it
  - Get victim to send an email with known content
- With known plaintext, easy to recover keystream
  - $C \oplus M = (M \oplus RC4(IV,key)) \oplus M = RC4(IV,key)$
- Even without knowing the plaintext, can exploit plaintext regularities to recover partial keystream
  - Plaintexts are not random: for example, IP packet structure is very regular
- Not a problem if the keystream is not re-used

Keystream Will Be Re-Used

- In WEP, repeated IV means repeated keystream
- Busy network will repeat IVs often
  - Many cards reset IV to 0 when re-booted, then increment by 1 $\Rightarrow$ expect re-use of low-value IVs
  - If IVs are chosen randomly, expect repetition in $O(2^{12})$ due to birthday paradox
- Recover keystream for each IV, store in a table
  - (KnownM $\oplus$ RC4(IV,key)) $\oplus$ KnownM = RC4(IV,key)
- Wait for IV to repeat, decrypt, enjoy plaintext
  - (M' $\oplus$ RC4(IV,key)) $\oplus$ RC4(IV,key) = M'

It Gets Worse

- Misuse of RC4 in WEP is a design flaw with no fix
  - Longer keys do not help!
    - The problem is re-use of IVs, their size is fixed (24 bits)
  - Attacks are passive and very difficult to detect
- Perfect target for the Fluhrer et al. attack on RC4
  - Attack requires known IVs of a special form
  - WEP sends IVs in plaintext
  - Generating IVs as counters or random numbers will produce enough "special" IVs in a matter of hours
- This results in key recovery (not just keystream)
  - Can decrypt even ciphertexts whose IV is unique

Fixing the Problem

- Extensible Authentication Protocol (EAP)
  - Developers can choose their own authentication method
    - Passwords (Cisco EAP-LEAP), public-key certificates (Microsoft EAP-TLS), passwords OR certificates (PEAP), etc.
- 802.11i standard fixes 802.11b problems
  - Patch (TKIP): still RC4, but encrypts IVs and establishes new shared keys for every 10 KBytes transmitted
    - Use same network card, only upgrade firmware
    - Deprecated by the Wi-Fi alliance
  - Long-term: AES in CCMP mode, 128-bit keys, 48-bit IVs
    - Block cipher in a stream cipher-like mode