RC4

Šifra, která míchá karty

Jednou z nejpoužívanějších proudových šifer v internetové a komerční kryptografii je RC4. Po sedm let zůstávala zahalena obchodním tajemstvím firmy RSA. Poté ji ale rozkryl (disassembloval) jeden hacker a popis zveřejnil na internetu. Z několika hledisek to je neobvyklá šifra. Seznámíme vás s její definicí a vlastnostmi.

Popis RC4 se poprvé objevil na internetu v poštovní konferenci „cypherpunks“ v roce 1994. Byl to krátký zdrojový kód, který sem zaslal anonymní hacker. Definice RC4, která do té doby byla chráněným obchodním tajemstvím společnosti **RSA Data Security Inc.** (dnes divize Security Dynamics Inc.), se prostřednictvím internetu rázem roznesla po celém světě. Ukázalo se, že z konstrukčního hlediska je RC4 zajímavá a neobvyklá šifra. Je řádově desetkrát rychlejší než DES a používá ji řada programů, zejména amerických. Například je použita v protokolu Secure Socket Layer 3.0 firmy Netscape, v Microsoft Office, v Oracle Secure SQL, ve Windows 2000 i jinde.

Komerční kryptografie je obvykle veřejná

Současným světovým trendem v komerční kryptografii je používání veřejných šifer. To je zcela v pořádku, protože tyto šifry používá široká veřejnost a zajišťují se jimi různé bezpečnostní služby (důvěrnost, autentizace, integrita aj.); proto je také správné, aby jejich kvalita mohla být veřejně posuzována. Takových šifer je sice početně málo, ale v komerční oblasti mají velmi silné procentuální zastoupení. Dá se dokonce říci, že se o jiných šifrách vlastně ani neví. RC4 byla výjimkou, která byla nasazena ve značné části softwaru, neboť za ní stála dostatečně silná americká společnost. Zajímavé je, že málokdo ji přitom podezíral, že by použila slabou šifru.

Je každá utajená šifra slabá?

Až do svého odhalení patřila šifra RC4 do třídy tzv. ***proprietárních algoritmů***. Některé zkušenosti s nepublikovanými slabými šiframi vedly k dosti rozšířené tendenci nedůvěřovat takovým šifrám a považovat je a priori za méně hodnotné. U šifry RC4 se však toto všeobecné mínění z neznámých příčin neuplatnilo. Je také dobré si uvědomit, že ve světě existuje velký počet dalších proprietárních šifer, o kterých nemáme ani tušení a které jsou také kvalitní. Jsou to utajované šifry používané v ozbrojených silách (vojsko, policie, rozvědka, kontrarozvědka), v bankovnictví a části průmyslu, ve státní správě, v diplomacii i ve vládě. Je jich mnohem více než šifer publikovaných. Jejich utajení přitom vůbec nepramení z obavy o kvalitu, ale je bezpečnostním opatřením, které zásadně znepříjemňuje život případným útočníkům na bezpečnostní systémy nebo šifrovaná data.
To byla ostatně i příčina utajení RC4. Měla chránit citlivá data zákazníků včetně čísel kreditních karet, privátních dokumentů ap. Pokud bychom postupovali podle zakořeněného zjednodušeného schématu „proprietární rovná se slabý“, RC4 by měla být slabou šifrou. Je tomu ale právě naopak.

RSA se zlobí

Krátce po zveřejnění údajného kódu RC4 bylo v téže poštovní konferenci potvrzeno, že se shoduje s výsledky šifry RC4 z oficiálního toolkitu společnosti RSA. RSA pak vydala prohlášení, že tento akt porušil právo, že je to zneužití internetu a že přijme opatření proti tomu, kdo by chtěl narušit duševní vlastnictví firmy. Asi se bála, aby šifru nezačali používat její konkurenti v komerčních produktech, protože nebyla v USA patentována.
K tomu ovšem nedošlo a nedošlo ani k rozšíření šifry mimo kontrolu RSA. Nebyl totiž důvod. Ve světě byla k dispozici celá řada jiných kvalitních a jako freeware právně zcela bezkonfliktních šifer. Ale i tak existuje právně nenapadnutelná cesta, jak ji použít i v USA. Stačí napsat, že se používá algoritmus se jménem třeba XRC4, který je *datově kompatibilní* s RC4 od RSA.

Popis RC4

RC4 je klasický symetrický algoritmus s tajným klíčem. Je to proudová šifra, kterou navrhl Ronald Rivest (RC znamená **R**ivest's **C**ipher), jeden z vynálezců algoritmu RSA a spoluzakladatel společnosti RSA DSI.
Vstupem RC4 je klíč o volitelné délce, teoreticky až 256 bajtů. Klíč inicializuje konečný automat, který pak generuje posloupnost bajtů hesla **h(0)**, **h(1)**, ... Při zašifrování se heslo „xoruje“ na otevřený text a při odšifrování na šifrový text, tedy: **št(i) = ot(i) xor h(i)**, i = 0, 1, ...
Základem konstrukce RC4 je princip podobný **míchání karet**. Mějme třeba 256 karet v nějakém základním zamíchání, které si označíme jako **karta(0), ..., karta(255)** a které vyložíme za sebou na stůl. (Na kartách mohou také být napsána čísla, s nimiž můžeme dělat různá

infotipy

O publikaci RC4:
http://www.lbbs.org/zmag/articles/chen.htm
Další podrobnosti a zdrojové kódy v C, ASM:
ftp://ftp.funet.fi/pub/crypt/cryptography/symmetric/rc4/
O útoku na RC4 v protokolu SSL:
http://pauillac.inria.fr/~doligez/ssl/

kouzla, ale o tom až později.) Každé pořádné míchání má být náhodné, předpokládejme tedy, že máme k dispozici 256 na kartách zcela nezávislých náhodných „míchacích“ (z množiny 0 až 255) čísel **r(i)**, i = 0, 1, ... 255. Často používaným míchacím principem je tento postup:

1. krok: vyměníme karty na pozici 0 a r(0);
2. krok: vyměníme karty na pozici 1 a r(1);
3. krok: vyměníme karty na pozici 2 a r(2);

...

256. krok: vyměníme karty na pozici 255 a r(255).

Takto vezmeme do ruky celkem 512 karet, takže každá v **průměru dvakrát** změní své místo. Protože čísla **r(i)** jsou náhodná, budou se mezi nimi vyskytovat některá čísla z množiny 0 ... 255 vícekrát, zatímco jiná vůbec. Některé karty se tedy budou přesunovat vícekrát, jiné jen jednou. Princip míchání zaručuje, že každá karta bude vzata alespoň jednou do ruky, ale kam a kolikrát se přesune, to záleží na celé posloupnosti **r.**

V extrémním případě může jedna karta změnit své místo i 256krát, tj. v každém kroku míchání. Pokud si pod kartami představíme čísla tak, že **karta(i) = i**, pak z počáteční identické permutace čísel (karet) 0 ... 255 máme na konci míchání náhodnou permutaci čísel 0 ... 255, jejíž prvky závisí na všech náhodných hodnotách **r**. Ideální princip pro šifru! RC4 tak také, jen nepatrně složitěji, vytváří svoji permutaci (substituční tabulku) **S** pro inicializaci generátoru hesla.

Karty míchá šifrovací klíč

Šifrovací klíč RC4 (uvažuje se zarovnaný na bajty) opakujeme tolikrát za sebou, až naplníme pole 256 bajtů **K(0)**, **K(1)**, ... , **K(255)**. Poté zvolíme identickou počáteční permutaci **S**, tj. **S(i) = i**, i = 0 ... 255, a promícháme ji prostřednictvím hodnot **K(i)**, které postupně učiníme ještě trochu složitějšími. Jestliže průběžný index označíme i = 0 ... 255, mícháme podle tohoto pseudokódu:

```
j = 0
for i = 0 to 255
{
    j = (j + S(i) + K(i)) mod 256
    S(i) <—> S(j)
}
```

V každém míchacím kroku se tedy vyměňují prvky permutace **S** na pozicích **i** a **j**. Index **i** je průběžný, zatímco „míchací index“ **j** závisí na klíči. Kdyby bylo ve vzorci použito jen **j = K(i)**, byl by to přesně případ míchání popsaný v předchozím odstavci. Nedosáhli bychom ale tak dobrého promíchání, protože posloupnost **K** není náhodná, ale naopak se v ní míchací bajty opakují!

Pro ilustraci zvolme čtyřbajtový klíč 50, 100, 131, 212. Konečným výsledkem míchání je posloupnost 50, 100, 131, 212, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ..., která příliš náhodně nevypadá. Aby se opakování v posloupnosti **K** eliminovalo, míchací index **j** závisí nejen na **K**, ale na všem, co se v každém předchozím kroku měnilo, tj. na všech (!) předchozích průběžných hodnotách **j**, **S** a **K**. Díky tomu míchací index **j** závisí na klíči velmi složitě, náhodně a periodičnost je tedy odstraněna.

Jak se generuje heslo

Po inicializační fázi se s mícháním pokračuje, ale tentokrát už každý krok produkuje jeden bajt hesla podle následujícího pseudokódu takto (všechna sčítání jsou v modulu 256):

```
j = 0, i = 0,
for t = 0 to N
{
    i = i + 1
    j = j + S(i)
    S(i) <—> S(j)
    h(t) = S(S(i) + S(j))
}
```

Délka klíče

RC4 se nejvíce používá s délkou klíče 40 nebo 128 bitů. Delší klíč je používán na území USA a kratší klíč pro export. Právě toto omezení způsobuje, že protokol SSL při ustavení šifrovaného kanálu u spojení mezi neamerickým klientem a americkým webem uměle sníží délku klíče na 40 bitů. Těchto 40 bitů se doplní veřejnou náhodnou informací vyměněnou mezi oběma stranami na počátku protokolu SSL a na tento řetězec se aplikuje hašovací funkce MD5 (viz též Chip 4/99, str. 44). Z hašovacího kódu se použije 88 bitů k doplnění původních 40, čímž je vytvořen 128bitový klíč pro RC4. Jeho efektivní délka ale zůstává 40 bitů.

Útok na RC4 v protokolu SSL

Vzhledem k obavám z nedostatečné bezpečnosti 40bitového klíče byla na internetu také zveřejněna výzva k rozluštění jedné reálné zprávy. Šlo o zachycení skutečné komunikace mezi klientem a webovým serverem pomocí protokolu SSL. Bylo v ní mj. zašifrováno i číslo kreditní karty. První výzva byla zveřejněna 14. 7. a druhá 19. 8. 1995. U první trvalo zjištění klíče osm dní, u druhé 32 hodin. Byl

přitom použit jenom triviální útok hrubou silou, kdy byl prostě zkoušen jeden 40bitový klíč za druhým, ale celá akce vzbudila na internetu velký rozruch. Šifra RC4 tím však nijak poškozena nebyla. Pro zajímavost dodejme, že RC4 dostala generální povolení od NSA k vývozu (nemuselo se žádat na každý případ zvlášť), pokud délka klíče bude redukována na 40 bitů.

Nyní je už jeden a tři čtvrtě roku možné vyvážet šifry s délkou klíče 56 bitů, ale kvůli kompatibilitě to mnoho producentů softwaru nevyužilo. Průkopníkem je, zdá se, Microsoft, který do Windows 2000 implementoval bezpečnostní protokol Kerberos. Ten dříve používal DES, ale nyní má nově všude jako přednastavenou šifru definovánu právě RC4 s 56 bity klíče (pro Američany 128).

Kryptografická kvalita

Téměř všechny dosud publikované proudové šifry jsou založeny na lineárních posuvných registrech se zpětnou vazbou a následnou nelineární funkcí nebo nepravidelným nelineárním krokováním registrů. Tyto konstrukce mají výhodu, že je u nich teoreticky dobře zvládnuta otázka periodičnosti, lineární složitosti a statistických vlastností. To se ale o RC4 říci nedá. Ta je založena na principu konečného automatu, přičemž geniální myšlenka míchání pochází z roku 1965 od MacLarena a Marsaglii, kteří na ní jako první založili generátor pseudonáhodných čísel (blíže viz Chip 6/98, str. 46).

Vnitřní stav automatu lze charakterizovat indexy **i** a **j** a obsahem permutace **S**. Automat přechází z jednoho stavu do druhého a na základě každého vnitřního stavu se vypočte jeden bajt hesla.

Všech jeho možných stavů je
256*256*(1*2*3* ... *256),
což je maximální možná délka periody hesla. Toto číslo je přibližně rovné 2^M, kde M = 1700.

RC4 je automatem, v němž se z následujícího stavu dá přejít do stavu předcházejícího jednoznačným způsobem. O těchto automatech víme, že jejich stavy mohou tvořit seskupení různě dlouhých cyklů o délkách 1, 2, ... , 2^M, přičemž všechny cykly jsou stejně pravděpodobné. V tomto seskupení se typicky objevuje jeden velký cyklus s délkou kolem 2^{M-1} a zbytek tvoří menší cykly různých délek. Průměrná perioda je z pravděpodobnostního hlediska 2^{M-1}, což u RC4 je dostatečně velké číslo 2^{1699}. O dalších vlastnostech toho není mnoho známo.

Výsledky teoretického výzkumu

Zatím byly publikovány dvě zajímavé práce, které zkoumaly RC4 analyticky. Jde o Goličův příspěvek „Linear Statistical Weakness of Alleged RC4 keystream Generator“, přednesený na konferenci Eurocrypt´97, a o příspěvek vědců Knudsena, Meiera, Preenela, Rijnmena a Verdoolaege s názvem „Analysis Methods for (Alleged) RC4“, který zazněl na konferenci Asiacrypt´98.

V první práci známý bělehradský specialista na proudové šifry zkoumá možnost lineární aproximace produkovaného hesla. Jeho výsledek lze zjednodušeně zformulovat následovně. Vezměme vždy nejnižší bit každého bajtu **h(i)** a označme jej **z(i)**. Posloupnost **z** binárně dvakrát zderivujme, čímž obdržíme posloupnost **d(i)** ≡ **z(i) xor z(i + 2)**. U posloupnosti **d** bylo zjištěno, že její prvky **d(i)** mají tendenci být spíše jednička než nula, a to s pravděpodobností 0,5 + 0,000000447. Přitom se tato korelace dá detekovat po cca 10^{12} bitech. Je to velmi hezký teoretický, ale, jak jistě vidíte, prakticky nepoužitelný výsledek.

Druhý příspěvek se snaží odvodit počáteční permutaci **S** na základě znalosti poměrně krátkého úseku hesla (což je oprávněný předpoklad) – a autoři vyvinuli algoritmus, který to umí. Je rychlejší než postupné zkoušení všech možných permutací, neboť má složitost blízkou odmocnině všech možných permutací. I když je to velký teoretický úspěch, pro praktický útok to znamená stále ještě příliš velký počet operací (přes 10^{234}).

Závěr

RC4 je zajímavá, neobvyklá a výjimečná šifra – jedna z mála velice rozšířených proprietárních šifer, která zůstávala po sedm let tajná. Vymyká se také z obecně oblíbeného omylu, že proprietární šifry musí být slabé. Kupodivu na teoretickém poli bylo o ní dosud publikováno málo výsledků. Možná je tomu tak právě proto, že každému kryptologovi je na první pohled jasné, že si na ní vyláme zuby.

Vlastimil Klíma (*vklima@decros.cz*)