Bezpečnost RSA

RSA v novém světle

(3)

Minule jsme poznali, jak snadno lze kvalitní šifru RSA „nabourat“ postranním kanálem. Jak se tomu ale bránit, když postranních kanálů je obrovské množství (a spousta se jich teprve zrodí v mozcích budoucích kryptoanalytiků)? Ukážeme si, že obranný manévr existuje – a dokonce obecný.

V tomto dílu si dovolíme vysvětlit námi navrhovanou metodu obrany proti obecným útokům založeným na postranních kanálech. Metoda je v řadě případů poměrně snadno prakticky implementovatelná, a přitom podle teoretického rozboru poskytuje užitečné obecné výsledky.

Z výkladu o postranních kanálech (dostupný on-line na [1], speciálně viz [2]) víme, že **postranním kanálem** nazýváme každý nežádoucí způsob výměny informací mezi kryptografickým modulem a jeho okolím. Takto volná definice naznačuje, o jak široké oblasti vlastně hovoříme; méně už je zřejmé, co si pod tímto pojmem představit konkrétně. V **kryptoanalýze** to moc nevadí, neboť zde většinou pracujeme naráz jen s úzce specifickými druhy kanálů, kde se už s jejich přesným popisem nějak dokážeme vypořádat (mnohdy jej ani nepotřebujeme a těžiště leží v popisu metod *analýzy a útoku* – podrobněji viz výše uvedené odkazy).

Jiná situace je však v **kryptografii**. Zde s ohledem na to, že chceme vytvořit konstrukci odolnou vůči současným i budoucím druhům útoků, potřebujeme pro postranní kanál nějaký přesnější a zároveň dostatečně obecný model. Pro naše účely dobře vyhoví model analogický k *obecnému modelu diskrétního kanálu*, který se již řadu let úspěšně používá v teorii informace.

Model postranního kanálu

V rámci našeho modelu si jako X nazveme diskrétní náhodnou veličinu značící vstupující informaci a jako Y diskrétní náhodnou veličinu značící informaci vystupující z daného postranního kanálu. U obou veličin předpokládáme konečný obor hodnot, přičemž uvažujeme jen ty hodnoty, kterých tyto veličiny nabývají s nenulovou pravděpodobností. Tento předpoklad je oprávněný, neboť zdrojem veličiny X je zde vždy nějaký počítač, který odpovídá konečnému automatu, a veličina Y je zase vyhodnocována nějakým konečným automatem útočníka. (Předesíláme, že tímto modelem postranního kanálu nechceme pokrýt kanály založené na kvantové teorii informace.)

Vlastní kanál popíšeme maticí, kterou vidíme v pravé části obrázku 1. Tato matice má tvar **SC = ($P_{i,j}$)**, kde $P_{i,j} = P[Y = y_j \mid X = x_i]$. Vidíme, že jednotlivé řádky matice SC (označení od výrazu *Side Channel*) odpovídají příslušným vstupním hodnotám a jednotlivé sloupce zase korespondují s hodnotami výstupu. Konkrétní prvek matice pak odpovídá podmíněné pravděpodobnosti, že na výstupu se objeví hodnota y_j za předpokladu, že na vstupu je hodnota x_i. Matici SC budeme také nazývat **kanálovou maticí**.

> Postranním kanálem nazýváme každý nežádoucí způsob výměny informací mezi kryptografickým modulem a jeho okolím.

Vzhledem k tomu, že pracujeme s pravděpodobnostmi, lze pro prvky kanálové matice poměrně snadno odvodit následující základní vztahy:

$$\Sigma_{(j)} P_{i,j} = \Sigma_{(j)} P[Y = y_j \mid X = x_i] = 1 \quad (1)$$

$$\Sigma_{(i)} \; \Sigma_{(j)} P[X = x_i] * P_{i,j} = \Sigma_{(i)} \; \Sigma_{(j)} P[X = x_i] * P[Y = y_j \mid X = x_i] = 1 \quad (2)$$

První z uvedených rovnic tvrdí, že každá ze vstupních hodnot se s jistotou nějak projeví na výstupu. Druhá zobecňuje první a říká, že pokud se „něco" objeví na vstupu postranního kanálu, potom se vždy objeví „něco" na jeho výstupu. Tyto interpretace jsou celkem srozumitelné a vyhovují všem známým druhům postranních kanálů.

Vzhledem k tomu, že máme absolutní volnost v přiřazení konkrétních významů jednotlivým hodnotám vstupní a výstupní veličiny (čili ve vytváření sémantiky nad daným kanálem), můžeme tímto modelem pokrýt i takové druhy kanálů, které by na první pohled některou z podmínek nesplňovaly. Například pokud by existovaly vstupní hodnoty, na které kanál nijak (měřitelně) nereaguje, potom bychom tento stav netečnosti shrnuli pod vybranou hodnotu veličiny Y a dále ji považovali za ordinérní druh reakce („Žádná odpověď – také odpověď", jak praví lidová moudrost).

Zkusme teď určit přenosové vlastnosti postranního kanálu. Použijeme konstrukci založenou na vyjádření množství informace o veličině X, která je obsažena ve veličině Y. V anglické literatuře se pro tento pojem používá výraz *mutual information*, tedy „vzájemná informace", my zde budeme ještě používat termín **informační přenos** (či přenos informace). Označíme jej **I(X; Y)** a definujeme takto:

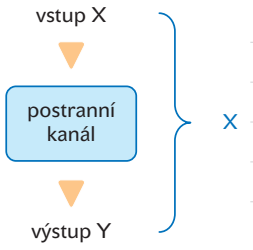
$$I(X; Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) = I(Y; X) \quad (3)$$

Všimněme si, že množství informace o veličině X obsažené ve veličině Y je stejné jako množství informace o veličině Y obsažené ve veličině X. (Právě tato vlastnost vedla k volbě názvu „vzájemná informace".)

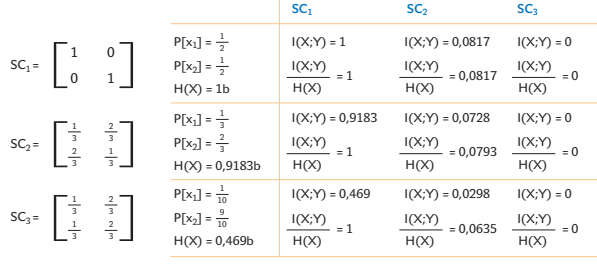
Výrazem H(X) zde rozumíme **entropii** veličiny X, výraz H(X | Y) popisuje podmíněnou entropii veličiny X za předpokladu znalosti hodnoty veličiny Y. Analogicky chápeme i výrazy H(Y) a H(Y | X). V dalším textu budeme u čtenáře předpokládat alespoň základní povědomí o konceptu entropie a k výkladu tohoto pojmu se proto nevracíme.

O informačním přenosu

Pro lepší přehled si výpočet přenosu I(X; Y) naznačíme v jeho významných krocích. Mějme dáno rozdělení vstupní veličiny X jako distribuční funkci **P[X = x_i]** a matici postranního kanálu **SC = ($P_{i,j}$)** typu [m, n]. To znamená, že veličina X může nabývat (s nenulovou pravděpodobností) celkem **m** různých hodnot, z nichž každá se může projevit jako **n** různých hodnot výstupní veličiny Y. Předpokládejme



Obr. 1. Popis postranního kanálu pomocí kanálové matice



Obr. 2. Příklady výpočtu informačního přenosu

útočníka, který sleduje výstupní veličinu Y. Naším úkolem bude určit, jak velké množství informace o vstupní veličině X takový útočník může získat.

Nejprve si na základě matice SC určíme distribuční funkci veličiny Y jako $P[Y = y_j]$. Pak můžeme psát

$$P[Y = y_j] = \Sigma_{(i)} P_{i,j} * P[X = x_i] \quad (4)$$

Na základě získané distribuční funkce již snadno určíme entropii H(Y) jako

$$H(Y) = \Sigma_{(j)} P[Y = y_j]*\log_2(P[Y = y_j]^{-1}) \quad (5)$$

Zde sčítáme přes všechny nenulové hodnoty distribuční funkce $P[Y = y_j]$. Dále pokračujeme ve výpočtu podmíněné entropie H(Y | X):

$$H(Y \mid X) = \Sigma_{(i)} P[X = x_i]*H(Y \mid X = x_i), \quad (6)$$

kde

$$H(Y \mid X = x_i) = \Sigma_{(j)} P_{i,j}*\log_2(P_{i,j}^{-1}) \quad (7)$$

Opět sčítáme přes všechny nenulové hodnoty $P_{i,j}$ a $H(Y \mid X = x_i)$. Pro snazší pochopení těchto vztahů připomeňme, že $P_{i,j} = P[Y = y_j \mid X = x_i]$. Nyní již zbývá jen dosadit do rovnice (3), kterou použijeme ve tvaru $I(X; Y) = H(Y) - H(Y \mid X)$.

Z uvedeného výpočtu vidíme, že výsledný informační přenos je závislý nejen na vlastnostech kanálu jako takového (ty zachycuje matice SC), ale i na rozdělení vstupní veličiny X. Konkrétně sem tato závislost vstupuje prostřednictvím rovnic (4) a (6). Podotkněme, že pokud bychom zvolili alternativní způsob výpočtu hodnoty I(X; Y), této závislosti bychom se nezbavili (jak vidíme z nutnosti výpočtu H(X)).

Smířit se s tímto poznatkem nám pomůže fyzikální příměr. Například v oblasti přenosu analogových signálů platí známé pravidlo o nutnosti vzájemného přizpůsobení impedance vysílače (zdroje informace) a vedení (přenosového kanálu). Námi pozorovaná závislost informačního přenosu na rozdělení veličiny X je vlastně analogií k tomuto impedančnímu přizpůsobení. Zjištění, že kvalita přenosu informací nezávisí jen na samotném kanálu, ale také na jeho sladění s vysílačem, tak máme potvrzeno i z „reálného" světa.

Pro větší názornost jsme na obrázku 2 vypsali kanálové matice pro tři konkrétní postranní kanály. Všechny jsou typu [2, 2], takže předpokládáme vstupní a výstupní veličiny nabývající nejvýše dvou různých hodnot. Připojená tabulka uvádí informační přenosy jednotlivých kanálů v závislosti na rozdělení vstupních hodnot.

Nulový informační přenos

Z obrázku 2 je patrné, že nejlepších přenosových výsledků dosahuje postranní kanál reprezentovaný jednotkovou maticí. To není velké překvapení, neboť takový kanál můžeme z fyzikálního hlediska považovat za ideální spojení vysílače s přijímačem. Kdybychom se zabývali „chtěným" přenosem informace, snažili bychom se dosáhnout právě takového stavu. Naším cílem však je najít takové podmínky, při nichž je informační přenos co nejhorší, tj. daným postranním kanálem prosakuje co nejméně informace.

Při pohledu na obrázek 2 dále vidíme, že nejhoršího přenosu dosahuje kanál, v jehož matici si jsou vektory všech řádků rovny. Lze dokázat, že takový kanál má bez ohledu na rozdělení vstupu vždy nulový informační přenos. Veličiny X a Y se za tohoto stavu chovají jako dvojice nezávislých náhodných veličin, takže $H(Y \mid X) = H(Y)$. Odtud pak přímo z rovnice (3) dostáváme, že $I(X; Y) = 0$. Ani toto nepřekvapuje, neboť se vlastně jedná o stav, kdy se všechny vstupní hodnoty projevují na výstupu statisticky identickým způsobem.

Zajištění nulového přenosu

Už tedy víme, jak by měla vypadat kanálová matice „neškodného" postranního kanálu – otázkou však zůstává, jak takovou matici vytvořit. Přímé ovlivnění fyzikálních vlastností daného kanálu je (s výjimkou použití dokonalého stínění) technologicky téměř vyloučeno – alespoň v obecném případě, a my se právě chceme na obecný případ zaměřit.

Na první pohled by se mohlo zdát, že jsme na tom podobně jako výzkumníci v oblasti teorie kódování – víme, jak by měla kanálová matice vypadat, ale nevíme, jak to prakticky zaručit. Zatímco v teorii kódování často nezbývá než tuto cestu zcela opustit a věnovat se pouze přizpůsobení vysílače (vhodným kódem), my zde jistou šanci máme. A máme ji právě proto, že chceme dosáhnout minimálního, a nikoliv maximálního přenosu.

Představme si, že sice nedokážeme měnit fyzikální vlastnosti daného kanálu, ale že máme možnost nechat zařízení před každou vyzářenou informací náhodně zvolit jeden z **r** postranních kanálů. Předpokládejme, že tato volba probíhá s rovnoměrným rozdělením a že r je velké. Formálně tato situace znamená, že místo jedné matice SC máme množinu matic $\{SC_1, SC_2, ..., SC_r\}$, z nichž se před každým odesláním informace do postranního kanálu náhodně vybere nějaká matice SC_i, podle níž bude daný přenos probíhat. Před příštím přenosem se volba matice opět opakuje.

Položme si teď otázku: Jak bude vypadat výsledná kanálová matice takto řízeného kanálu z pohledu útočníka? Opět není příliš těžké dokázat, že situace se mu bude jevit tak, jako by byl použit postranní kanál popsaný maticí

$$SC_C = r^{-1}\Sigma_{i=1}{}^{r} SC_i \quad (9)$$

V sumě je použit klasický maticový součet, násobení hodnotou r^{-1} představuje násobení matice skalárem. Zaměřme se nyní na chování hodnot v jednotlivých sloupcích výsledné matice SC_c (nazveme ji **maticí kanálové superpozice**). Zjednodušíme-li poněkud naše úvahy tím, že budeme odpovídající si hodnoty ve sloupcích matic SC_i považovat za hodnoty nezávislých náhodných veličin se stejným (po sloupcích) rozdělením, potom lze pro velká r podle zákona velkých čísel očekávat, že hodnoty ve sloupcích matice SC_c se budou blížit k určité střední hodnotě. Konkrétní číslo reprezentující tuto střední hodnotu zde pro nás není důležité. Důležité je, že vzdálenost mezi hodnotami ve sloupcových vektorech se bude s rostoucím r pravděpodobně zmenšovat, čímž se matice SC_c bude blížit tvaru, pro který dostáváme nulový informační přenos.

Tuto situaci názorně ilustruje obrázek 3, na kterém je zachycena hustota distribuční funkce náhodné veličiny vyjadřující poměrnou změnu v informačním přenosu. Graf (vykreslený programem *Mathematica 4*) byl získán tak, že se 1000krát náhodně vygenerovala sada 256 (tj. r = 256) kanálových matic typu [2, 2]. Pro každou sadu se vypočítala výsledná matice SC_c a vyhodnotila se poměrná změna přenosu pro každou matici ze sady jako $I_{SCc}(X; Y)/I_{SCi}(X; Y)$. Každá sada tak poskytla 256 údajů o relativní změně, takže celkem se v grafu zpracovalo 256 000 takových změn.

Pro tento ilustrační experiment jsme předpokládali, že vstupní veličina X má rovnoměrné rozdělení. Nechceme tvrdit, že přesně takto bude vypadat chování všech možných superpozic. Uvádíme to jen jako příklad popisující obecný trend relativní změny informačního přenosu, který plně podporuje námi odhadované chování celého systému. Tento trend říká, že ve většině případů dojde po superpozici k výraznému poklesu přenosu informace.

Parazitní vyzařování operací

Zbývá ještě vyřešit otázku, jak do systému zanést náhodnou volbu kanálové matice.

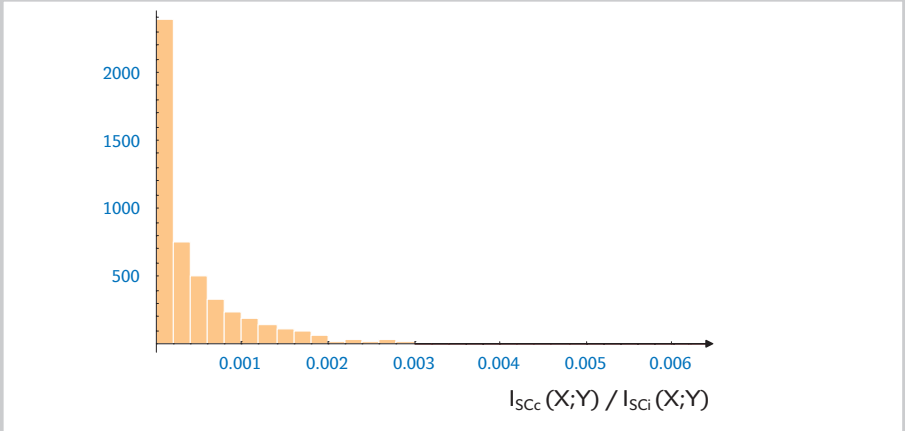
Chceme vytvořit konstrukci odolnou vůči současným i budoucím útokům.

Pro tento účel se zaměříme na konkrétní operace, které probíhají v námi sledovaném a zabezpečovaném modulu. **Parazitním vyzařováním** zvolené operace nazveme postranní kanál, který přenáší informaci o vstupních hodnotách této operace. Mějme například operaci f: A → Im(f). Potom parazitní vyzařování této funkce bude popsáno kanálovou maticí, jejíž počet řádků bude odpovídat počtu prvků z množiny A, které s nenulovou pravděpodobností vstupují do funkce f. Počet sloupců pak bude korespondovat s počtem různých znaků, které je možné pozorovat na výstupu daného postranního kanálu. (Pojem „znak" je v tomto kontextu třeba chápat velmi obecně.)

Funkce, o níž byla řeč, patří do kategorie unárních operací. Obecně si musíme představit n-ární operaci vystupující jako zobrazení **f: A_1 x A_2 x ... x A_n → Im(f)**. Předpokládejme dále, že jeden z argumentů o dostatečně velkém rozsahu hodnot (m) není pro výsledek operace sémanticky důležitý, takže jej můžeme použít k libovolnému účelu (konkrétně nechť to je a_n, nabývající r hodnot). Navíc víme, že n-ární operaci $f(a_1, a_2, ..., a_n)$ můžeme pro vybraný argument a_n popsat jako r (n-1)-árních operací $\{f_1(a_1, a_2, ..., a_{n-1}), f_2(a_1, a_2, ..., a_{n-1}), ..., f_m(a_1, a_2, ..., a_{n-1})\}$, kde hodnotu a_n dosazujeme vždy implicitně. Přitom každá z těchto funkcí má vlastní charakter parazitního vyzařování, který je popsán maticemi $\{SC_1, SC_2, ..., SC_r\}$. Volbou konkrétní hodnoty parametru a_n tak vlastně volíme konkrétní vyzařovací matici SC_i – a to je právě ten „trik", který jsme potřebovali.

Eliminace vyzařování prakticky

Ústřední myšlenkou popisované techniky je tedy zanesení náhodné volby některého z parametrů zabezpečované operace. Tento parametr musí mít dostatečně velký rozsah hodnot, aby se začal projevovat zákon velkých čísel pro výslednou kanálovou matici parazitního vyzařování, a zároveň nesmí ovlivnit sémantiku této operace v daném kontextu.



Obr. 3. Rozdělení relativní změny informačního přenosu pro superponovaný kanál

Představme si například, že potřebujeme ochránit součet dvou 16bitových (modulo 2^{16}) čísel a že máme k dispozici 32bitovou sčítačku. V takovém případě si můžeme za maskovací parametr zvolit obě horní (numericky významnější) poloviny vstupujících 32bitových slov, které naplníme náhodnými hodnotami. Do dolních polovin vstupních slov pak umístíme hodnoty, které chceme sečíst. Náhodné maskovací hodnoty nám zde suplují volbu jedné z 2^{32} kanálových matic, což by se mělo projevit výrazným poklesem nežádoucího informačního přenosu. Obdobně je možné maskovat operace násobení, logický součet, součin, nonekvivalenci a další.

Pár poznámek závěrem

Právě jsme si předvedli obecný model postranního kanálu a ukázali jsme si jeho souvislosti s parazitním vyzařováním operací probíhajících v kryptografických modulech. Zavedli jsme pojem informačního přenosu a odvodili jsme jeho závislost na matici postranního kanálu (SC). Na základě toho jsme prokázali kladný přínos techniky maskování citlivých operací náhodnou volbou sémanticky nedůležitých vstupních parametrů pro potlačení parazitního vyzařování těchto operací.

Je třeba upozornit, že navrhovaná technika má sloužit zejména jako preventivní doplňková ochrana. Jejím detailním rozborem jsme zde chtěli ukázat, že má svůj smysl, a že je tudíž vhodné věnovat jí při návrhu kryptografických modulů pozornost. Netvrdíme však, že tato technika je schopná nahradit ostatní protiopatření konstruovaná přímo proti konkrétním druhům útoků – na to je příliš obecná. V kombinaci s těmito protiopatřeními zato její obecnost pomáhá čelit dosud neznámým druhům útoků, u nichž může výrazně zbrzdit jejich dopad. Protože útoky se většinou zdokonalují postupně, mohou tak konstruktéři získat čas, aby na nově vzniklé útoky dokázali reagovat vývojem cílených intenzivních protiopatření.

Při odhadu síly konstruovaných mechanismů jsme vyšli důsledně z teorie informace. Alternativně se v kryptografii využívá přístup založený na teorii složitosti, která bere ohled na výpočetní sílu protivníka. (Díky tomu se používá častěji, neboť poskytuje v jistém smyslu reálnější pohled než přístup informační, který implicitně předpokládá útočníka disponujícího neomezeným výpočetním potenciálem.) Odtud plyne, že informační pohled je mnohem přísnější než složitostní, což nás opravňuje k domněnce, že výsledný návrh má dobré předpoklady v praxi obstát.

K této problematice se ještě jednou vrátíme. V příštím pokračování si ukážeme konkrétní využití popsané metody v návrhu procedury pro odšifrování zpráv pomocí RSA podle standardu PKCS#1.

Vlastimil Klíma, *autor@chip.cz*
Tomáš Rosa, *autor@chip.cz*

Literatura

[1] Archiv článků
http://www.decros.cz/bezpecnost/_kryptografie.html
[2] Rosa, T.: Kryptoanalýza s využitím postranních kanálů, Vojenská kryptografie IV, Sborník příspěvků, str. 113–156, 2001, dostupné v [1].