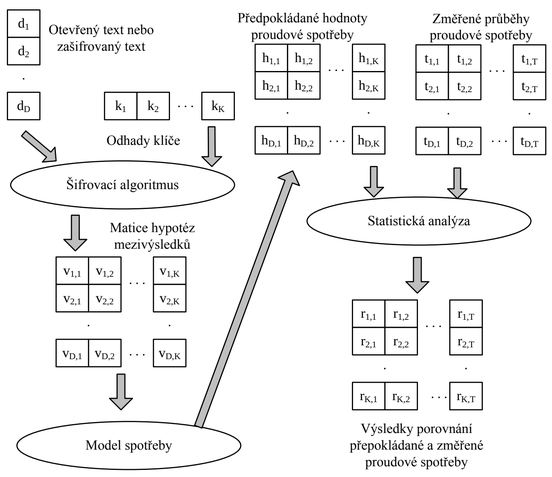
Diferenciální proudová analýza DPA

Zdeněk Martinásek

Vysoké učení technické v Brně, Ústav telekomunikací
Purkynova 118, Brno
martinasek@feec.vutbr.cz

1. Diferenciální proudová analýza DPA

V následujícím textu bude popsán detailněji postup získání tajného klíče metodou DPA. Tento postup je popsán v pěti krocích a využívají ho všechny útoky DPA (obr. 1).



Obrázek 1: *Blokový diagram znázorňující kroky* 3 *až* 5 *DPA útoku*

První krok: Volba mezivýsledku algoritmu
Prvním krokem DPA je volba mezivýsledku kryptografického algoritmu, který je vykonáván zařízením. Mezivýsledek musí být funkcí $f(d, k)$, kde d jsou známá nekonstantní data a k je malá část tajného klíče (např. první bajt). Ve většině případů útoku DPA d je otevřený text nebo šifrovaný text. Takto definovaný mezivýsledek může být použit k určení části tajného klíče k.
Druhý krok: Měření proudové spotřeby
Druhým krokem DPA útoku je měření výkonové spotřeby kryptografického zařízení při šifrování nebo dešifrování různých bloků dat D. Pro všechny operace šifrování či dešifrovaní potřebuje útočník znát hodnoty zpracovávaných dat d, které se podílí na výpočtu mezivýsledku zvoleného v prvním kroku. Hodnoty známých dat tvoří vektor $d = (d_1, \ldots, d_D)^{'}$, kde d_i označuje hodnotu i-tého kroku šifrování nebo dešifrování.
V průběhu každého tohoto kroku si útočník zaznamenává proudovou spotřebu. Průběhy proudové spotřeby korespondující k bloku dat d_i označíme $t_i^{'} = (t_{i,1}, \ldots, T_{i,T})$, kde T označuje délku naměřené proudové spotřeby. Útočník měří výkonovou spotřebu pro každý blok dat D naměřené průběhy můžou být zapsány maticově **T** o velikosti $D \times T$. Pro DPA útok je klíčové, aby naměřené proudové průběhy byly přesně zarovnané. To znamená že hodnoty pro jednotlivé sloupce t_j matice **T** musí odpovídat stejné operaci. K získání takto zarovnaných dat je nutná správná synchronizace s měřícím zařízením nebo je zapotřebí zarovnat data softwarově pomocí nalezení několika markantů (otisků v proudovém průběhu).
Třetí krok: Výpočet hypotetických mezivýsledků Dalším krokem útoku je výpočet hypotetických mezivýsledků pro všechny možné hodnoty klíče k. Všechny možnosti klíče lze zapsat jako vektor $k = (k_1, \ldots, k_K)$, kde K označuje celkový počet možných klíčů. V DPA jsou jednotlivé prvky vektoru k označovány za hypotézy klíče nebo odhady klíče. Z vektoru známých dat d a vektoru hypotéz všech klíčů může útočník jednoduše vypočítat hodnotu mezivýsledku $f(d, k)$ pro všechny šifrovací operace D a pro všechny hypotézy klíče K. Výsledkem výpočtu (1) je matice **V** o rozměrech $D \times K$.

$$v_{i,j} = f(d_i, k_j)\ i = 1, \ldots, D\ j = 1, \ldots, K \tag{1}$$

Sloupec j matice **V** obsahuje mezivýsledky, které byly vypočítány dle hypotéz klíče k_j. Je zřejmé, že jeden sloupec matice **V** obsahuje takové mezivýsledky které byli vypočítány v zařízení během operce šifrování a dešifrování. Jinými slovy jednotlivé sloupce matice **V** obsahují mezivýsledky pro všechny klíč, tedy i pro klíč který byl použit v zařízení. Tento index bude označen ck, tedy k_{ck} označuje hledaný tajný klíč. Cílem DPA je nalezení odpovídajícího sloupce, který byl zpracováván při operacích šifrování a dešifrování v zařízení a tedy nalezení k_{ck}.
Čtvrtý krok: Mapování hypotetických mezivýsledků na hodnoty proudové spotřeby
Čtvrtým krokem DPA útoku je namapování matice hypotetických mezivýsledků **V** na matici **H** reprezentující předpokládané hodnoty výkonové spotřeby. V tomto bodě se využívá simulace výkonové spotřeby kryptografického zařízení. Použitý model spotřeby přiřadí každému hypotetickému mezivýsledku $v_{i,j}$ předpokládanou hodnotu výkonové spotřeby $h_{i,j}$. Správnost výsledků simulace silně závisí na útočníkových znalostech o zařízení a činní DPA efektivnější. Mezi často používané modely přiřazení hodnot **V** na **H** patří model Hammingovy vzdálenosti a Hammingovy váhy.
Pátý krok: Porovnání hypotetických hodnot proudové spotřeby s naměřenými průběhy
V posledním kroku útočník porovná předpokládané hodnoty výkonové spotřeby závislé na odhadu klíče (hodnoty ve sloupci h_i matice **H**) se změřenými průběhy proudové spotřeby (hodnoty ve sloupci t_j matice **T**). Výsledkem je matice **R** velikosti $K \times T$, kde každý element $r_{i,j}$ představuje výsledek porovnání sloupců h_i a t_j. Porovnání je provedeno postupy popsanými v následujících kapitolách. Společná vlastností všech postupů je, že hodnota $r_{i,j}$ je větší pro lepší shodu sloupců h_i a t_j. Tajný klíč je určen na základě následujících poznatků.

2. Útok založený na korelačním koeficientu

Korelační koeficient (Correlation coefficient) patří k nejznámější metodě k určení lineární závislosti mezi dvěma náhodnými proměnnými. Proto je to také vhodná metoda pro provedení DPA útoku. Existuje velmi dobře definovaná teorie pro korelační koeficient, který může být použít k modelování statických vlastností DPA útoků. Korelační koeficient je definován pomocí kovariance vztahem:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{\sigma^2(X) \cdot \sigma^2(Y)}}. \tag{2}$$

Jedná se o bezrozměrnou veličinu a může nabývat hodnotu $-1 \leq \rho \leq 1$. Hodnota -1 korelačního koeficientu značí nepřímou závislost (změna v jedné skupině je provázena opačnou změnou ve skupině druhé). Hodnota 0 korelačního koeficientu značí, že mezi hodnotami obou skupin neexistuje žádná statisticky zjistitelná závislost. Jestliže korelační koeficient je roven 1, značí to přímou závislost, dokonalou korelaci mezi hodnotami obou skupin. Také ρ je většinou neznáme a je nutné hodnotu odhadnout. Tento odhad r je definován vztahem:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x}) \cdot (y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \cdot \sum_{i=1}^{n}(y_i - \overline{y})^2}}. \tag{3}$$

V DPA je korelační koeficient použit k určení lineární závislosti mezi sloupci h_i a t_j pro $i = 1, \ldots, K$ a $j = 1, \ldots, T$. Výsledkem je matice **R** obsahující korelační koeficienty. Označíme každou hodnotu jako $r_{i,j}$ na základě elementů D ze sloupců h_i a t_j. Použijeme li předchozí definici korelačního koeficientu můžeme vztah 3 vyjádřit:

$$r_{i,j} = \frac{\sum_{d=1}^{D}(h_{d,i} - \overline{h_i}) \cdot (t_{d,j} - \overline{t_j})}{\sqrt{\sum_{d=1}^{D}(h_{d,i} - \overline{h_i})^2 \cdot \sum_{d=1}^{D}(t_{d,j} - \overline{t_j})^2}}, \tag{4}$$

kde h_i a $\overline{t_j}$ označují průměrné hodnoty sloupců h_i a t_j.

3. Útok založený na rozdílu středních hodnot

Základem statistické metody založené na rozdílu středních hodnot (Difference of mean) je srovnání dvou skupin naměřených hodnot (distribucí) výpočtem rozdílu středních hodnot těchto skupin. Tato metoda používá jiný způsob k určení mezi sloupci matice **H** a **T**. Útočník vytvoří binární matici **H**, která rozdělí naměřené proudové průběhy do dvou skupin. Posloupnost nul a jedniček v každém sloupci **H** je funkcí vstupních dat d a odhadů klíče k_i. Za účelem ověření zda odhad klíče k_i je správný útočník může rozdělit matici **T** na dva soubory řádků (tzn. dvě sady proudových spotřeb podle h_i). První soubor obsahuje ty řádky **T** index odpovídá pozici nul ve vektoru h_i. Druhý soubor obsahuje zbylé řádky **T**. Následně útočník vypočítá průměr řádků. Vektor $m_{0i}^{'}$ značí průměr řádků prvního souboru a $m_{1i}^{'}$ označuje průměr druhého souboru. Odhad klíče k_i je správný pokud existuje výrazný rozdíl mezi $m_{0i}^{'}$ a $m_{1i}^{'}$. Rozdíl mezi $m_{0i}^{'}$ a $m_{1i}^{'}$ indikuje vztah mezi h_{ck} a některým ze sloupců **T**. Stejně tak jako v předchozím případě tato diference označuje časový okamžik kdy je mezivýsledky odpovídající h_{ck} zpracovávány. V jiných okamžicích je diference mezi vektory rovna nule. Výsledkem útoku je tedy matice **R**, kde každý řádek odpovídá rozdílu mezi vektory $m_{0i}^{'}$ a $m_{1i}^{'}$ pro jeden odhad klíče. Rovnice výpočtu **R** je dána vztahem:

$$m_{1i,j}^{'} = \frac{1}{n_{1i}} \cdot \sum_{l=1}^{n} h_{l,i} \cdot t_{l,j}, \tag{5}$$

$$m_{0i,j}^{'} = \frac{1}{n_{0i}} \cdot \sum_{l=1}^{n} (1 - h_{l,i}) \cdot t_{l,j}, \tag{6}$$

$$n_{1,i} = \sum_{l=1}^{n} h_{l,i}, \tag{7}$$

$$n_{0i} = \sum_{l=1}^{n} (1 - h_{l,i}), \tag{8}$$

$$\mathbf{R} = \mathbf{M_1} - \mathbf{M_0} \tag{9}$$

kde n značí počet řádků matice **H** (tzn. počet naměřených proudových spotřeb).

4. Útok založený na vzdálenosti středních hodnot

Tato metoda založená na vzdálenosti středních hodnot (Distance of Means) je vylepšení předchozí metody, protože bere v úvahu směrodatné odchylky. Metoda je založena na běžně používaném testu k porovnání rovnosti středních hodnot dvou různých rozdělení. Útok využívající tuto metodu rozdělí matici **T** na dvě skupiny řádků pro každý odhad klíče stejně tak jak v předchozí kapitole 3. Rozdíl od předchozí metody spočívá v tom, že střední hodnoty jsou porovnány podle testu vzdálenosti středních hodnot. Prvky matice **R** jsou vypočítány dle následujícího vztahu:

$$r_{i,j} = \frac{m_{1i,j} - m_{0i,j}}{s_{i,j}}, \tag{10}$$

kde $s_{i,j}$ je směrodatná odchylka pro rozdělení dvou skupin.

EUROPEAN UNION
MINISTRY OF EDUCATION, YOUTH AND SPORTS
OP Education for Competitiveness