

Analýza AES white-box schémat pomocí útoku postranním kanálem

Autor: Jakub Klemsa

Školitel: prof. RNDr. Václav Matyáš, M.Sc., Ph.D.

Katedra matematiky

FJFI, ČVUT

23. března 2016

1. Ingredience

- Advanced Encryption Standard (AES)
- Od black-box k white-box modelu
- White-box AES (WBAES)
- Útok postranním kanálem (SCA)

2. Útok na WBAES

- Využití nástrojů SCA k útoku na WBAES
- Nové terče

3. Výsledky

- Reprodukce výsledků Bos et al.
- Výsledky použití všech 255 terčů
- Útok „naslepo“

4. Budoucí práce

Advanced Encryption Standard (AES)

Symetrická bloková šifra

- standard vydaný NIST [6, 2001]
- dodnes považována za bezpečnou

128-bitová varianta

- 10 rund, 4 elementární operace
- mezivýsledky zranitelné

Advanced Encryption Standard (AES)

Symetrická bloková šifra

- standard vydaný NIST [6, 2001]
- dodnes považována za bezpečnou

128-bitová varianta

- 10 rund, 4 elementární operace
- **mezivýsledky zranitelné**

Black-box model

Útočník má černou skříňku, která

- uchovává náhodný AES klíč
- zašifruje vloženou zprávu
- neunikne **žádná** další informace
 - mezivýsledky, doba šifrování, ...

Snaha útočníka: **získat klíč**

- uhodnutý klíč snadno ověří

Black-box model

Útočník má černou skříňku, která

- uchovává náhodný AES klíč
- zašifruje vloženou zprávu
- neunikne **žádná** další informace
 - mezivýsledky, doba šifrování, ...

Snaha útočníka: **získat klíč**

- uhodnutý klíč snadno ověří

Gray-box model

Blíže realitě

- z reálného hardwaru informace uniká
 - spotřeba energie, EM záření, ...

⇒ útok postranním kanálem (dále)

White-box model

Útočník vykonává šifrování

- vidí mezivýsledky
- může měnit hodnoty, instrukce, ...
- klíč musí zůstat utajen

Pozorování

Odolnost k white-box \Rightarrow odolnost ke gray-box.

White-box AES (WBAES)

Poprvé představeno Chow et al. [3, 2002]

- algebraický útok (Billet et al. [1, 2004])

Skrývá mezivýsledky

- plně tabulková implementace
- tabulky obklopené náhodnými bijekcemi
- vhodně se vyruší \Rightarrow zachová funkcionalitu

White-box AES (WBAES)

Poprvé představeno Chow et al. [3, 2002]

- algebraický útok (Billet et al. [1, 2004])

Skrývá mezivýsledky

- plně tabulková implementace
- tabulky obklopené náhodnými bijekcemi
- vhodně se vyruší \Rightarrow zachová funkcionálnitu

Útok postranním kanálem (SCA) I

Kombinuje

- zranitelnost mezivýsledků
- únik informace v gray-box modelu (nazveme *stopa*)

Celá řada SCA

- přímé pozorování klíče (RSA)
- hádání klíče **po částech** a hledání náznaků mezivýsledků ve stopách

Útok postranním kanálem (SCA) I

Kombinuje

- zranitelnost mezivýsledků
- únik informace v gray-box modelu (nazveme *stopa*)

Celá řada SCA

- přímé pozorování klíče (RSA)
- hádání klíče **po částech** a hledání náznaků mezivýsledků ve stopách

Útok postranním kanálem (SCA) II

Příklad útoku proti AES

- prochází hodnoty i -tého bytu klíče
 - dělí stopy podle j -tého bitu očekávaného mezivýsledku
 - nazveme *terč* (máme jich 8)
 - liší se podle zprávy
- max. rozdíl středních hodnot \sim správný klíč

Útok postranním kanálem (SCA) II

Příklad útoku proti AES

- prochází hodnoty i -tého bytu klíče
 - dělí stopy podle j -tého bitu očekávaného mezivýsledku
 - nazveme *terč* (máme jich 8)
 - liší se podle zprávy
- max. rozdíl středních hodnot \sim správný klíč

Útok postranním kanálem (SCA) II

Příklad útoku proti AES

- prochází hodnoty i -tého bytu klíče
 - dělí stopy podle j -tého bitu očekávaného mezivýsledku
 - nazveme *terč* (máme jich 8)
 - liší se podle zprávy
- max. rozdíl středních hodnot \sim správný klíč

Útok postranním kanálem (SCA) II

Příklad útoku proti AES

- prochází hodnoty i -tého bytu klíče
 - dělí stopy podle j -tého bitu očekávaného mezivýsledku
 - nazveme *terč* (máme jich 8)
 - liší se podle zprávy
- max. rozdíl středních hodnot \sim správný klíč

Využití nástrojů SCA k útoku na WBAES

Pozorování

Odolnost k white-box \Rightarrow odolnost ke gray-box (tj. SCA).

Bos et al. [2, čvc. 2015]: SCA na white-box implementace

- zlomili všechny veřejně dostupné
- jen jedno tabulkové AES (Klinec [4])
 - není zřejmé, kudy informace uniká
- paměťové stopy
 - adresy čtení/zápisu
 - obsah paměti

Využití nástrojů SCA k útoku na WBAES

Pozorování

Odolnost k white-box \Rightarrow odolnost ke gray-box (tj. SCA).

Bos et al. [2, čvc. 2015]: SCA na white-box implementace

- zlomili všechny veřejně dostupné
- jen jedno tabulkové AES (Klinec [4])
 - není zřejmé, kudy informace uniká
- paměťové stopy
 - adresy čtení/zápisu
 - obsah paměti

Využití nástrojů SCA k útoku na WBAES

Pozorování

Odolnost k white-box \Rightarrow odolnost ke gray-box (tj. SCA).

Bos et al. [2, čvc. 2015]: SCA na white-box implementace

- zlomili všechny veřejně dostupné
- jen jedno tabulkové AES (Klinec [4])
 - není zřejmé, kudy informace uniká
- paměťové stopy
 - adresy čtení/zápisu
 - obsah paměti

Využití nástrojů SCA k útoku na WBAES

Pozorování

Odolnost k white-box \Rightarrow odolnost ke gray-box (tj. SCA).

Bos et al. [2, čvc. 2015]: SCA na white-box implementace

- zlomili všechny veřejně dostupné
- jen jedno tabulkové AES (Klinec [4])
 - není zřejmé, kudy informace uniká
- paměťové stopy
 - adresy čtení/zápisu
 - obsah paměti

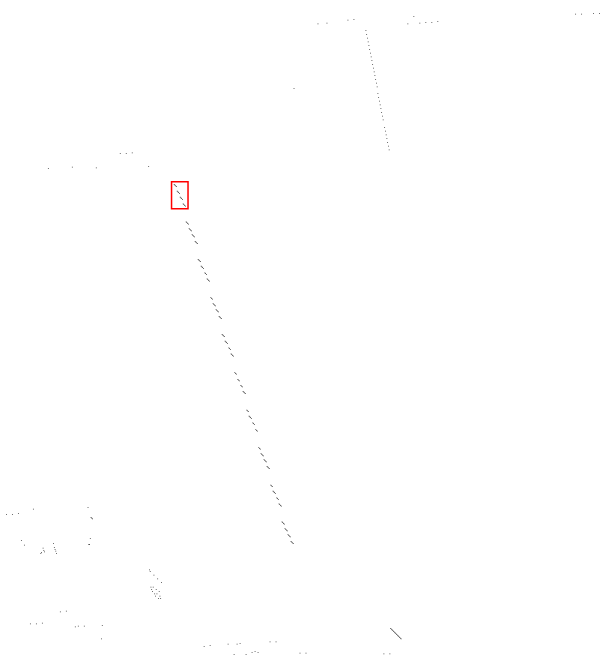
Využití nástrojů SCA k útoku na WBAES

Pozorování

Odolnost k white-box \Rightarrow odolnost ke gray-box (tj. SCA).

Bos et al. [2, čvc. 2015]: SCA na white-box implementace

- zlomili všechny veřejně dostupné
- jen jedno tabulkové AES (Klinec [4])
 - není zřejmé, kudy informace uniká
- paměťové stopy
 - adresy čtení/zápisu
 - obsah paměti



Obrázek : Paměťová stopa běhu AES (část). První runda vyznačena červeně.

Nové terče

WBAES obklopuje mezivýsledky mj. náhodným lineárním zobrazením

- původní terče \sim násobení mezivýsledku e.g. $(0, 0, 1, 0, 0, 0, 0, 0)$
- nové terče \sim násobení lib. nenulovým vektorem

\Rightarrow 255 nových terčů

Reprodukce výsledků Bos et al.

Bos et al. – útok na WBAES

- 8 původních + 8 dalších terčů (jiný přístup)
- všech 16 funguje
- reprodukuje k porovnání s novými terči

Byte	Původní terče							
	1.	2.	3.	4.	5.	6.	7.	8.
1.	■	55	90	■	149	207	224	■
2.	248	218	239	244	247	■	251	247
3.	■	212	■	25	230	■	99	■
4.	■	252	226	247	■	255	241	252
5.	247	104	■	225	229	■	225	249
6.	252	255	■	241	242	■	4	255
7.	47	233	■	228	■	■	■	■
8.	■	253	253	■	255	251	1	■
9.	224	196	231	249	253	238	■	253
10.	■	■	255	245	255	■	234	■
11.	245	■	250	■	190	255	236	■
12.	254	255	■	255	■	■	■	■
13.	241	■	254	190	160	193	■	■
14.	235	■	254	■	255	2	■	255
15.	■	■	246	195	255	■	246	155
16.	252	255	254	■	251	245	235	■

Tabulka : Pořadí správného kandidáta, ■ ~ 0. Útok s použitím 1024 stop. Průměrně 2.9 terčů z 8 uspěje (36%).

Byte	Dalších 8 terčů							
	1.	2.	3.	4.	5.	6.	7.	8.
1.	207	■	4	252	253	252	■	■
2.	233	255	252	■	216	255	255	■
3.	254	209	■	■	254	225	247	189
4.	37	■	251	■	■	252	231	242
5.	244	■	250	231	134	79	214	223
6.	■	253	255	254	■	■	■	2
7.	■	248	187	255	209	■	184	227
8.	■	255	255	242	234	253	■	255
9.	227	156	237	243	229	232	■	■
10.	■	158	1	■	253	■	■	■
11.	248	■	241	254	251	45	255	1
12.	■	251	254	255	236	255	■	254
13.	205	4	191	30	■	■	240	255
14.	■	231	246	■	248	253	■	■
15.	221	250	1	■	223	■	1	225
16.	255	■	229	254	■	255	254	253

Tabulka : Pořadí správného kandidáta, ■ ~ 0. Útok s použitím 1024 stop. Průměrně 2.4 terčů z 8 uspěje (30%).

Výsledky použití všech 255 terčů

Nagenerováno 8 instancí WBAES tabulek

- 255 terčů pro každý ze 16 bytů klíče a každou z 8 instancí
 - 32 640 útoků
 - doba běhu desítky hodin
- průměrná úspěšnost 29%
 - silní kandidáti – rozdíl na druhého $> 10\%$
 - úsp. 25%

Výsledky použití všech 255 terčů

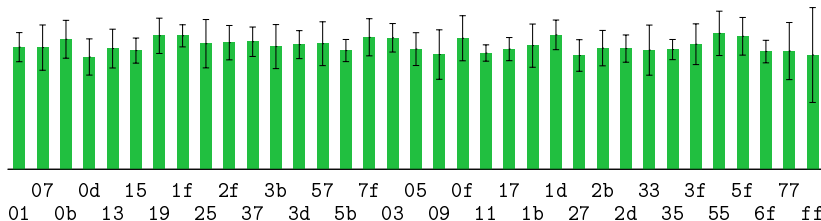
Nagenerováno 8 instancí WBAES tabulek

- 255 terčů pro každý ze 16 bytů klíče a každou z 8 instancí
 - 32 640 útoků
 - doba běhu desítky hodin
- průměrná úspěšnost 29%
 - silní kandidáti – rozdíl na druhého $> 10\%$
 - úsp. 25%

Rozložení úspěšnosti terčů I

Mnoho terčů, málo měření

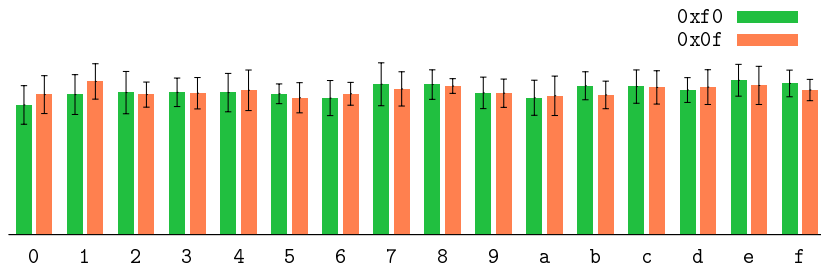
- shlukování terčů – cyklické rotace vektoru
- hlubší význam v konstrukci



Obrázek : Úspěšnost shluků terčů, hexadecimální zápis reprezentanta.

Rozložení úspěšnosti terčů II

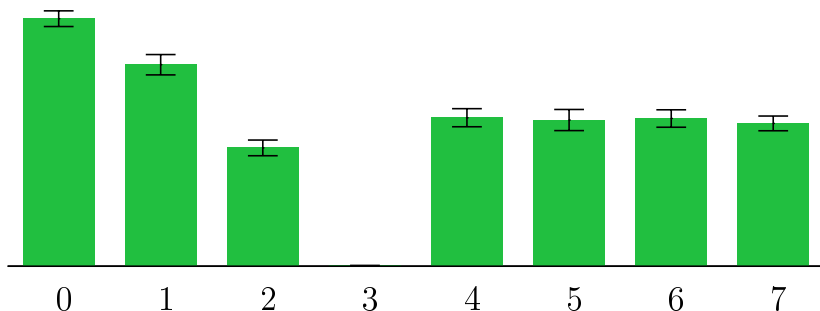
Fixované 4 bity vektoru (maskami 0xf0, 0x0f)



Obrázek : Úspěšnost shluků terčů, hexadecimální zápis fixovaných bitů.

Úspěšnost podle pozice ve stopě

Místo úniku – max. rozdíl středních hodnot stop



Obrázek : Průměrná úspěšnost a její směrodatná odchylka napříč instancemi podle pozice bitu v rámci bytu ve stopě.

Útok „naslepo“

Znali jsme klíč, úskalí útoku „naslepo“

- falešní kandidáti
 - rekordní rozdíl na druhého téměř 35%
- průměrný rozdíl správných 34%

Návrh

- použití menšího počtu stop a více terčů
- sčítání relativních rozdílů na druhého
 - hranice pro započítání (10%)
 - hranice pro skončení (75%)

Útok „naslepo“

Znali jsme klíč, úskalí útoku „naslepo“

- falešní kandidáti
 - rekordní rozdíl na druhého téměř 35%
- průměrný rozdíl správných 34%

Návrh

- použití menšího počtu stop a více terčů
- sčítání relativních rozdílů na druhého
 - hranice pro započítání (10%)
 - hranice pro skončení (75%)

Budoucí práce

Nedořešené otázky

- ze které tabulky WBAES dochází k úniku
 - 3 nefungující způsoby
 - 2 způsoby s debuggerem
 - útok na indexy tabulek
- jak útok teoreticky podložit
 - proč se správný kandidát propadá na poslední místa
- proč z pozice 3 ve stopě neuniká informace
- ...

Literatura I



Olivier Billet, Henri Gilbert, and Charaf Ech-Chatbi.

Cryptanalysis of a white box aes implementation.

In *Selected Areas in Cryptography*, pages 227–240. Springer, 2004.



Joppe W Bos, Charles Hubain, Wil Michiels, and Philippe Teuwen.

Differential computation analysis: Hiding your white-box designs is not enough.

Technical report, Technical report, Cryptology ePrint Archive, Report 2015/753. <http://eprint.iacr.org/2015/753>, 2015.



Stanley Chow, Philip Eisen, Harold Johnson, and Paul C Van Oorschot.

White-box cryptography and an aes implementation.

In *Selected Areas in Cryptography*, pages 250–270. Springer, 2002.

Literatura II



D. Klinec.

Whitebox-crypto-AES.

Git repository.

<https://github.com/ph4r05/Whitebox-crypto-AES>.



Dušan Klinec.

White-box attack resistant cryptography.

2013.



NIST FIPS PUB.

197: Advanced encryption standard (aes).

Federal Information Processing Standards Publication,

197:441–0311, 2001.