VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

ANALÝZA HLAVNÍCH KOMPONENT V PROUDOVÉ ANALÝZE

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

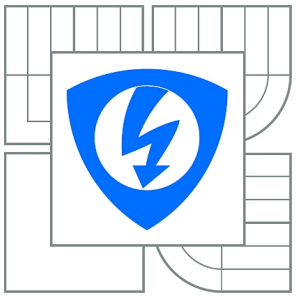
AUTOR PRÁCE
AUTHOR

FRANTIŠEK JEDLIČKA

BRNO 2015

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

ANALÝZA HLAVNÍCH KOMPONENT V PROUDOVÉ ANALÝZE

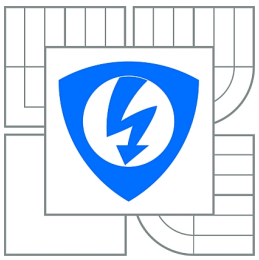
PRINCIPAL COMPONENT ANALYSIS IN POWER ANALYSIS

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE — FRANTIŠEK JEDLIČKA
AUTHOR

VEDOUCÍ PRÁCE — Ing. ZDENĚK MARTINÁSEK, Ph.D.
SUPERVISOR

BRNO 2015



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky a komunikačních technologií

Ústav telekomunikací

Bakalářská práce

bakalářský studijní obor
Teleinformatika

Student: František Jedlička ***ID:*** 152658
Ročník: 3 ***Akademický rok:*** 2014/2015

NÁZEV TÉMATU:

Analýza hlavních komponent v proudové analýze

POKYNY PRO VYPRACOVÁNÍ:

V rámci bakalářské práce se seznamte s problematikou kryptoanalýzy proudovým postranním kanálem. V teoretické části práce se zaměřte na použití analýzy hlavních komponent v proudové analýze a na obdobné používané techniky předzpracování dat. Vypracujte přehled současného stavu problematiky. V praktické části práce realizujte implementaci metody hlavních komponent (PCA) na proudové průběhy algoritmu AES a následně proveďte proudovou analýzu (př. CPA). Porovnejte dosažené výsledky, časovou náročnost a paměťovou náročnost útoku s analýzou hlavních komponent a bez ní.

DOPORUČENÁ LITERATURA:

[1] Mangard, S.; Oswald, E.; Popp, T.: Power Analysis Attacks: Revealing the Secrets of Smart Cards (Advances in Information Security). Secaucus, NJ, USA:Springer-Verlag New York, Inc., 2007, ISBN 0387308571.
[2] Kocher, P. C.; Jaffe, J.; Jun, B.: Differential Power Analysis. In CRYPTO '99: Proceedings of the 19th Annual International Cryptology Conference onAdvances in Cryptology, London, UK: Springer-Verlag, 1999, ISBN 3-540-66347-9, s. 388–397.

Termín zadání: 9.2.2015 ***Termín odevzdání:*** 2.6.2015

Vedoucí práce: Ing. Zdeněk Martinásek, Ph.D.
Konzultanti bakalářské práce:

doc. Ing. Jiří Mišurec, CSc.
Předseda oborové rady

UPOZORNĚNÍ:

ABSTRAKT

Tato práce se zabývá využitím analýzy hlavních komponent v kryptoanalýze proudovým postranním kanálem. Nejdříve je v práci rozebrána problematika kryptoanalýzy, kryptoanalýzy proudovým postranním kanálem, metody analýzy hlavních komponent a interpretace obdržených proudových průběhů provedené diferenciální proudové analýzy na kryptografickém modulu algoritmu AES. Praktická část obsahuje provedení vlastní analýzy hlavních komponent na obdržená data a následný pokus o diferenciální proudovou analýzu takto upravených dat.

KLÍČOVÁ SLOVA

Analýza hlavních komponent(PCA), proudová analýza, DPA, postranní kanál, DTW, ICA

ABSTRACT

This thesis deals with using principal component analysis in cryptanalysis by power side chanel. At first in this thesis is discussed cryptanalysis, cryptanalysis by power side chanel, principal component analysis method and interpretation received power consumption from performed differential power analysis on cryptographic device with AES algorithm. Practical part contain execution of own principal component analysis on received data and following try of differential power analysis thus adjusted data.

KEYWORDS

Principal component analysis, power analysis, DPA, side chanel, DTW, ICA

JEDLIČKA, František *Analýza hlavních komponent v proudové analýze*: bakalářská práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2015. 38 s. Vedoucí práce byl Ing. Zdeněk Martinásek, Ph.D.

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Analýza hlavních komponent v proudové analýze“ jsem vypracoval(a) samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor(ka) uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil(a) autorská práva třetích osob, zejména jsem nezasáhl(a) nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom(a) následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno ....

podpis autora(-ky)

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu diplomové práce panu Ing. Zdeňkovi Martináskovi, Ph.D. za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Brno ....

podpis autora(-ky)

Faculty of Electrical Engineering
and Communication
Brno University of Technology
Purkynova 118, CZ-61200 Brno
Czech Republic
http://www.six.feec.vutbr.cz

PODĚKOVÁNÍ

Výzkum popsaný v této bakalářské práci byl realizován v laboratořích podpořených z projektu SIX; registrační číslo CZ.1.05/2.1.00/03.0072, operační program Výzkum a vývoj pro inovace.

Brno . . . . . .

. . . . . .
podpis autora(-ky)

EVROPSKÁ UNIE
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ
INVESTICE DO VAŠÍ BUDOUCNOSTI

OP Výzkum a vývoj
pro inovace

OBSAH

SEZNAM OBRÁZKŮ

SEZNAM TABULEK

ÚVOD

V dnešní době se klade velký důraz na zabezpečení kryptografických modulů jak po stránce softwarové tak i po stránce hardwarové. Před proudovou analýzou lze do značné míry ochránit kryptografický modul například odstraněním závislosti právě prováděných šifrovacích operacích na čase, tímto postupem lze de-synchronizovat měření proudové spotřeby a tím znesnadnit proudovou analýzu.

Cílem této práce je provedení analýzy hlavních komponent na rozsynchronizované proudové průběhy naměřené na kryptografickém modulu s implementovaným algoritmem AES s použitým mikroprocesorem ATmege16.

V práci je teoretický rozbor proudové analýzy zaměřený na diferenciální proudovou analýzu s užitím metody korelačních koeficientů. Dále detailní popis analýzy hlavních komponent s možnostmi použití PCA v prostředí Matlab a popis možných funkcí které lze aplikovat.

V praktické části je uvedení a popis proudových průběhů, prvotní pokus o diferenciální proudovou analýzu. Dále následuje uskutečněná aplikace analýzy hlavních komponent na proudových průbězích v prostředí Matlabu, a opětovný útok diferenciální proudovou analýzou.

1 ÚTOKY POSTRANNÍMI KANÁLY

V této kapitole bych chtěl vysvětlit, co je to kryptografie, kryptoanalýza, kryptografický modul, jaké jsou základní útoky na kryptografický modul, co je to útok postranním kanálem a jaké druhy postranních kanálů jsou dnes již známy. Dále bych chtěl trochu podrobněji uvést problematiku proudové analýzy.

1.1 Základní pojmy

Kryptografie

Kryptografie je věda zabývající se utajováním zpráv a cenných informací a zajišťující jejich bezpečnost a utajení šifrováním.

Kryptoanalýza

Kryptoanalýza je věda zabývající se snahou odhalit tajné informace zabezpečené šifrováním, v podstatě opak kryptografie.

Kryptografický modul

K zajištění bezpečnosti kryptografického systému využíváme následující služby: důvěryhodnost, autentičnost, integrita a nepopiratelnost. Tyto služby zajišťuje kryptografický modul, který je v podstatě fyzickou realizací daného algoritmu či protokolu. Může se jednat o realizaci hardwarovou, softwarovou a kombinovanou. Lze tedy říci, že kryptografický modul je zařízení, uvnitř kterého probíhá celý proces šifrování a dešifrování. Modul pracuje s citlivými daty například s tajným klíčem.

Kryptografický modul používá pro komunikaci s okolním přesně definovaných vstupů a výstupů pro šifrování a dešifrování. [5] [7] [9]

Útoky postranním kanálem

Kryptografický modul kromě komunikace přes předem definované vstupy a výstupy v reálném světě komunikuje s prostředím i jinými nežádoucími způsoby. Během chodu kryptografického modulu probíhají uvnitř zařízení určité operace které jsou závislé na právě zpracovávaných datech. Tyto operace se díky fyzické konstrukci daného modulu projevují nežádoucím způsobem a tak komunikují s okolním prostředím. Jedná se například o proudovou spotřebu, o vyzařované teplo a elektromagnetické napětí nebo závislosti určité operace dobu na čase. Této komunikaci se říká postranní kanál. [5] [9]

1.2 Druhy postranních kanálů

Zde je uveden základní přehled nejznámějších postranních kanálů, které jsou dnes známy. Hodlám uvést: časový, elektromagnetický, optický, chybový, akustický a proudový postranní kanál. Informace použité ve zbytku této kapitoly vychází z prací [8] [5] [7] [9].

Časový postranní kanál

U tohoto typu útoku se předpokládá, že každá operace s daty trvá určitý čas závislý na právě zpracovávaných datech. Operace může být například nějaká instrukce.

Elektromagnetický postranní kanál

Elektromagnetický postranní kanál spočívá ve vyzařování nekonstantního elektromagnetického pole do okolí v závislosti na právě zpracovávaných datech. Toto pole může útočník lehce změřit pomocí elektromagnetických sond. Princip tohoto útoku je dost podobný proudovému postrannímu kanálu, který bude vysvětlen níže.

Optický postranní kanál

Tranzistory ze kterých se skládají integrované obvody vně kryptografických modulů při přetnutí svých stavů vyzařují fotony. Tyto fotony lze pomocí speciálních zařízeních snímat a jejich následná analýza může vést k odhalení senzitivních informací. Tato metoda je poměrně nová a její provedení je značně nákladné.

Chybový postranní kanál

Jde o využití komunikace kryptografického modulu s okolím během chybného stavu. Chybová hlášení může útočník využít ve svůj prospěch uvedením kryptografického modulu do chybového stavu a jejich sledováním. Například může měnit napájecí napětí nebo na vstup šifrování přivádět data o špatném formátu.

Akustický postranní kanál

Jedná se asi o nejstarší postranní kanál. Nejčastěji se odposlouchávají stisky jednotlivých tlačítek klávesnice napadaného zařízení, které se pak analyzují. Dále lze tento kanál využít u odposlouchání mikroprocesoru a zaznamenávání jednotlivých prováděných operací.

1.3 Proudový postranní kanál

Předpokládejme, že kryptografický modul během procesu šifrování a dešifrování odebírá proud ze zdroje závisle na právě prováděných operacích, na zpracovávaných datech a na hodnotě šifrovacího klíče. Právě tohoto se využívá u kryptoanalýzy proudovým postranním kanálem (dále jen proudová analýza). [9]

Celkovou proudovou spotřebu kryptografického modulu můžeme vyjádřit jako součet odebíraného proudu ze všech jednotlivých buněk integrovaného obvodu. Tyto buňky, například invertor, jsou složeny z tranzistorů, které jsou většinou založeny na technologii CMOS. Procesor obsažený v kryptografickém modulu, s implementovaným šifrovacím algoritmem, se skládá z několika milionů takovýchto tranzistorů. Proudová spotřeba je tedy závislá na počtu a uspořádání těchto elementárních buněk uvnitř modulu [5]. Tranzistory CMOS jsou obvykle napájeny konstantním napětím, U_{DD}, dále celkový okamžitý proud je $i_{DD}(t)$. Pro úplnost průměrnou výkonovou spotřebu za čas T lze dostat vztahem

$$P_{obv} = \frac{1}{T}\int_0^T P_{obv}(t)dt = \frac{U_{DD}}{T}\int_0^T i_{DD}(t)dt \tag{1.1}$$

Tuto celkovou výkonovou spotřebu můžeme obecně rozdělit na statickou a dynamickou. Statická spotřeba P_{stat} je odebírána obvodem při stabilních stavech a dynamická spotřeba P_{dyn} je spotřeba, která nastane pokud se změní stav obvodu (dojde k přechodu). Celková spotřeba se pak rovná $P_{stat} + P_{dyn}$, všechny spotřeby pro možné stavy jsou v následující tabulce

Tab. 1.1: Proudová spotřeba přechodů

Přechod	**Výkonová spotřeba**	**Typ spotřeby**
0→0	P_{00}	statická
0→1	P_{01}	statická+dynamická
1→0	P_{10}	statická+dynamická
1→1	P_{11}	statická

Platí-li podmínka $P_{00} \approx P_{11} << P_{01}, P_{10}$, lze z tabulky 1.1 usoudit, že dynamická výkonová spotřeba je závislá na právě zpracovávaných datech. [5] [7] [9]

1.3.1 Jednoduchá proudová analýza

Pro přehled zde uvedu základní popis a rozdělení jednoduché proudové analýzy (*Simple power analysis* – SPA). Základní myšlenka jednoduché proudové analýzy je vliv proudové spotřeby kryptografického modulu na právě probíhajících šifrovacích algoritmech jak bylo popsáno výše. Výhodou této metody je, že stačí naměřit jen pár proudových průběhů, výjimečně stačí na úspěšný útok jen jeden změřený proudový průběh. V případě analýzy z více naměřených proudových průběhů(*single-shot* SPA) získáváme výhodu zprůměrování naměřených průběhů a tím odstranění šumu, na rozdíl od jediného průběhu (*multiple-shot* SPA). Nevýhodou metody je potřeba detailně znát kryptografický modul a šifrovací algoritmus. Pro přehled zde uvedu základní rozdělení SPA:

- Přímá interpretace proudových průběhů
- Analýza s využitím šablon
- Kolizní útok

Toto rozdělení už dále nebudu rozebírat, bylo by to nad rámec této práce. [5] [9]

1.3.2 Diferenciální proudová analýza

Diferenciální proudová analýza (*Differencial power analysis* – DPA) se provádí obdobně jako SPA. Na rozdíl od SPA útočník nepotřebuje znát podrobně kryptografický algoritmus, za to potřebuje změřit daleko více proudových průběhů pro různá šifrovaná data. Lze tedy říct, že výhodou DPA oproti SPA je jednoduchost a nízký nárok na znalosti o kryptografickém modulu. Naopak nevýhodou je potřeba naměřit poměrně hodně proudových průběhů, což přináší problémy jak s delší dobou trvání měření proudových průběhů tak i při zpracování objemného souboru dat. Další velkou výhodou DPA je možnost získat tajný klíč i z dat s poměrně velkou hodnotou šumu. [5] [9]

Existuje několik typů DPA lišící se jen v posledním kroku analýzy, známé jsou:

- Analýza s využitím šablon
- Útok založený na korelačním koeficientu
- Rozdíl středních hodnot
- Vzdálenost středních hodnot

Útok založený na korelačním koeficientu

Jak bylo uvedeno výše, existuje hned několik možností při výběru útoku DPA, všechny mají shodný postup, jen v posledním kroku se zvolí jiná metoda, podle které se daná analýza jmenuje.

Prvním krokem je volba mezivýsledku. Mezivýsledek kryptografického algoritmu se nejčastěji volí do místa nějaké nelineární operace v datech. Funkci mezivýsledku můžeme vyjádřit funkcí $f = (d, k)$, kde d jsou známá vstupní data různá pro každé měření proudového průběhu a k může být část šifrovacího klíče, například předpoklad prvního bajtu. [5] [9]

Druhým krokem je vlastní měření proudové spotřeby, proudových průběhů. Do zařízení vstupují bloky dat D různé pro každé měření, tyto data tvoří vektory dat d, která jsou známa. Pro každý vektor dat si posléze útočník změří proudovou spotřebu. Výsledná matice naměřených proudových spotřeb má poté rozměry $D \times T$, kde T odpovídá délce naměřené proudové spotřeby, závislá na době trvání šifrovacího algoritmu. Je nutné, aby byla data správně synchronizována, prakticky to znamená, aby data ve stejném sloupečku výstupní matice odpovídala stejné operaci. Bez tohoto zarovnání by DPA nebyla úspěšná. [5] [9]

Třetí krok, výpočet teoretických mezivýsledků. V tomto kroku se užije vybraný mezivýsledek z prvního kroku a vypočítá se funkce $f = (d, k)$. Je potřeba spočítat veškeré možné hypotetické průběhy pro všechny možnosti klíče z vektoru k a z vektorů dat d pro všechny šifrovací operace D. Výsledkem je matice V. [5] [9]

Čtvrtým krokem je mapování mezivýsledků do hodnot hypotetické proudové spotřeby. V tomto kroku se užije nějaký model proudové spotřeby, například Hammingovy váhy pro simulaci spotřeby z matice V v závislosti na zpracování dat D na výslednou matici teoretické proudové spotřeby H. Úspěšnost tohoto kroku závisí na útočníkově znalostech o daném zařízení. [5] [9]

Pátý krok je porovnání vypočtených teoretických hodnot s hodnotami naměřené proudové spotřeby. Princip této analýzy spočívá v předpokladu, že proudové průběhy budou v jistý časový okamžik závislé na vypočteném mezivýsledku. Vybraná metoda pomocí korelačního koeficientu je metoda k určení lineární závislosti mezi daty, je definována obecným vzorcem:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{\sigma^2(X) * \sigma^2(Y)}}, \tag{1.2}$$

kde $Cov(X, Y)$ je výpočet kovariance dle 2.1 a σ^2 je výpočet rozptylu dle 3.3, viz níže. Výstupem celého DPA útoku je pak matice korelací pro všechny teoretické odhady hodnot klíče s naměřenými průběhy. Počet řádků odpovídá počtu bitů klíče a počet sloupců je počet vzorků naměřených průběhů. Výsledkem korelace je hodnota od -1 do 1, princip tohoto výsledku je vysvětlen na kovarianci v 2.1. [5] [9]

2 ANALÝZA HLAVNÍCH KOMPONENT

Analýza hlavních komponent (Principal comonent analysis – PCA) je metoda předzpracování dat. Převážně se používá k potlačení šumu, nebo redukci dimenze dat. V podstatě se jedná o lineární metodu transformace původních proměnných na nové.

PCA se běžně používá ke snížení dimenze dat, jinými slovy zmenší se počet znaků a to za předpokladu že nedojde k velké ztrátě informace. Lze tuto metodu využít u DPA průběhů obdobným způsobem. Cílem je upravit nebo vybrat data tak, aby byla použitelná pro analýzu, snížila se hodnota šumu a nedošlo k příliš velké ztrátě užitečné informace. Obdobnými analýzami se zabývám v kapitole 3 [1] [13] [14].

2.1 Popis PCA

PCA můžeme popsat následujícími kroky:

Získání dat

Nejdříve musíme získat data. Naše data jsou již naměřené proudové průběhy.[8]

Normalizace dat

Dalším krokem je upravit data tak, aby byla středově centrovaná tj. průměr každé dimenze byl roven 0. Toho lze dosáhnout odečtením průměru dimenze od každého jejího prvku, takže pokud máme soubor dat o n dimenzích, musíme od každého prvku jednotlivé dimenze odečíst průměr dané dimenze, tedy $n_{i_j} - \overline{n_i}$ kde n_{i_j} je j-tý prvek i-tého rozměru a $\overline{n_i}$ je aritmetický průměr i-tého rozměru. Takto upravená data se nazývají normalizovaná. [13]

Výpočet kovarianční matice

V tomto kroku se musí vypočítat kovarianční matice z normalizovaných dat. Kovariance se dá popsat jako míra vzájemné vazby mezi dvěma dimenzemi. Pro kovarianci existuje obecný vzoreček.

$$Cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)}, \tag{2.1}$$

kde X a Y představují jednotlivé rozměry a n počet jejich prvků. Výsledek kovariance může nabývat hodnot od -1 do 1. Je-li výsledek kladný, mezi daty dimenzí existuje vzájemná kladná vazba tj. se vzrůstající hodnotou prvků jedné dimenze vzrůstá také hodnota prvků druhé. Pokud je výsledkem záporná hodnota, existuje záporná

vazba, a je-li výsledek 0 mezi daty je buď velmi slabá nebo žádná vazba. Pokud bude výsledkem kovariance 1, data jsou stejná. [13]

Pro úplnost zde uvedu výpočet rozptylu který udává střední kvadratickou odchylku.

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{(n-1)} \tag{2.2}$$

Porovnáním obou rovnic můžeme vidět, že kovariance jen jedné dimenze je rozptyl. Teď můžeme vypočítat kovarianční matici. Pro naše n rozměrná data je potřeba vypočítat kovarianci mezi všemi dimenzemi. Budeme tedy mít n^2 výpočtů tj. matici o rozměrech $n * n$. Matice obsahující například 3 dimenze bude vypadat následovně:

$$C = \begin{pmatrix} Cov(n_1, n_1) & Cov(n_1, n_2) & Cov(n_1, n_3) \\ Cov(n_2, n_1) & Cov(n_2, n_2) & Cov(n_2, n_3) \\ Cov(n_3, n_1) & Cov(n_3, n_2) & Cov(n_3, n_3) \end{pmatrix}$$

Zde si můžeme všimnout, že kovariance jsou postupně prováděny podle indexů řádků a sloupů matice, takže například na druhém řádku, třetím sloupci máme výpočet kovariance mezi druhým a třetím rozměrem. Vzhledem k tomu, že výsledek kovariance mezi jedním a druhým rozměrem je stejný jako mezi druhým a prvním, bude výsledná matice symetrická kolem hlavní diagonály, a vzhledem k tomu že výsledek kovariance jen mezi jedním rozměrem je rozptyl, hlavní diagonála bude obsahovat rozptyly jednotlivých dimenzí. [13] [14]

Výpočet vlastního vektoru a vlastního čísla

Vlastní vektor matice označuje vektor, jehož směr se při vynásobení dané matice tímto vektorem nezmění, změní se pouze velikost. Vlastní vektor lze popsat následujícím vztahem:

$$A * x = \lambda * x,$$

kde A je čtvercová matice, x je vlastní vektor matice A a λ je vlastní číslo tohoto vektoru. Takže každému vlastnímu vektoru náleží jedno vlastní číslo.

Vlastní vektor může mít pouze čtvercová matice, a ne každá čtvercová matice má vlastní vektor. Pokud máme čtvercovou n-rozměrnou matici symetrickou podél hlavní diagonály můžeme předpokládat, že lze vypočíst n vlastních vektorů a vlastních čísel. Pro PCA je důležité mít vlastní vektory v jednotkovém tvaru tj. velikost vektoru je rovna 1. [13]

Výběr vlastních vektorů

Následující krok je poměrně jednoduchý. Nyní musíme naši matici vlastních vektorů seřadit sestupně podle hodnot vlastních čísel, přičemž vlastní vektor s nejvyšší hodnotou vlastního čísla je hlavní komponenta.

Z takto seřazené matice vektorů nyní vybereme několik p prvních vektorů, hlavních komponent, které mají největší vypovídací schopnost o celém datovém souboru. Tyto vektory stejným způsobem seřadíme do nové matice a to: jednotlivé vektory budou ve sloupečcích seřazeny sestupně dle jejich hodnot vlastních čísel. Tato matice se nazývá feature vektor. [13] [1]

Výpočet nových dat

Konečným krokem je výpočet nového datového souboru. Ve chvíli kdy máme vybrány hlavní komponenty a z nich zformovaný feature vektor, nová data dostaneme:

$$NováData = FeatureVector^T \times NormalizovanáData^T,$$

kde symbol T značí transponovanou matici, podle pravidel pro operace s maticemi lze tedy počítat také:

$$NováData{=}(FeatureVector \times NormalizovanáData)^T.$$

Takto jsme získali z původních dat data nová vyjádřená pomocí vlastních vektorů, hlavních komponent. Jelikož jsou tyto vektory ortogonální tj. jsou si navzájem kolmé, dochází tak k redukci dimenze.

Můžeme také z těchto dat dostat data původní:

$$PůvodníData^T = (FeatureVector^T \times NormalizovanáData) + Průměry.$$

Tímto vzorečkem dostaneme přesná původní data jen pokud se rozhodneme ponechat si ve feature vektoru všechny hlavní komponenty. Zajímavé ale je, že tento vzoreček můžeme použít i pro redukovaný počet vlastních vektorů. [1]

3 OBDOBNÉ METODY PŘEDZPRACOVÁNÍ DAT POUŽÍVANÉ V PROUDOVÉ ANALÝZE

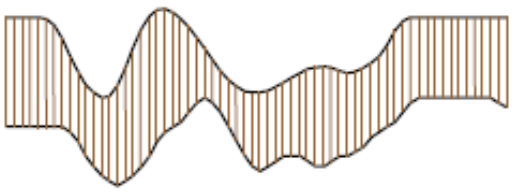
V této kapitole budu rozebírat nebo poukazovat na metody předzpracování signálu, které lze použít na úpravu proudových průběhů před proudovou analýzou. Některé tyto metody lze použít na prezentované proudové průběhy, a některé řeší trochu odlišné problémy s nezarovnanými a zašumělými daty.

Příklady používaných metod v proudové analýze

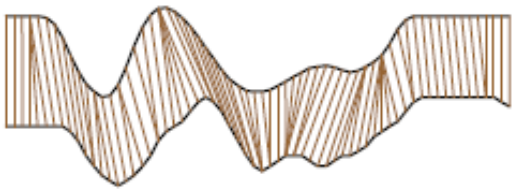
Obdobné metody používané v kryptografii na úpravu dat před vykonáním proudové analýzy mohou být například analýza nezávislých komponent, používaná k redukci šumu v datech se smíšenými signály. Dále shluková analýza, sloužící k rozčlenění podobných signálů do shluků, které nejsou podobné se signály v ostatních shlucích. Faktorová analýza zachycuje závislosti různých pozorování (signálů), pomocí několika faktorů. Metoda dynamického borcení časové osy, která upravuje vzdálenost bodů signálu v čase. Aplikaci ICA řeší práce [2], použitím metody DTW na proudovou analýzu se zabývá např. práce [3].

3.1 Dynamické borcení časové osy

Metoda Dynamického borcení časové osy (The dynamic time warping – DTW) se používá pro průběhy nesynchronizované na časové ose, například se používá u analýzy lidské řeči k vyhledávání klíčových slov. Metoda DTW měří vzdálenosti mezi dvěma význačnými body v čase a poté měří jejich vzdálenost. Z obrázků 3.1 a 3.2 si můžeme udělat představu o principu DTW (převzato z [3]).



Obr. 3.1: Obvyklá vzdálenost



Obr. 3.2: Zborcená vzdálenost

Deformace (borcení) časové osy se provádí na základě vypočítané cesty, která zarovná data pod sebe na základě signálu s nejmenší vzdáleností. Toto porovnání probíhá mezi časovými osami pouze dvou průběhů, je tedy pro tuto metodu požadován referenční signál [3].

Vypočítaná cesta vzdáleností (warp path) je v podstatě seznam indexů porovnávaných průběhů ukazujících, který prvek odpovídá prvku druhého průběhu k druhému prvku. Můžeme definovat tuto cestu F následující rovnicí

$$F = (c(1), c(2), ..., c(K)), \tag{3.1}$$

kde X a Y jsou porovnávané signály, takže platí $c(k) = (x(k), y(k))$. Pomocí tohoto formálního vyjádření si nyní můžeme definovat podmínky pro počítání vzdáleností jako:

Monotónnost: $x(k-1) \leq x(k), y(k-1) \leq y(k)$

Kontinuita: $x(k) - x(k-1) \leq 1, y(k) - y(k-1) \leq 1$

Hraniční podmínky: $x(1) = y(1) = 1, x(K) = T, y(K) = T$

Kde T je počet vzorků signálů X a Y. Dále pak warp path může být určena ohraničena:

$$T \leq K < 2T$$

Podmínka monotónnosti značí, nemožnost vzorky posouvat zpět na časové ose. Kontinuita vylučuje vynechání některých vzorků a hraniční podmínka předpokládá, že signály budou začínat i končit na stejném místě.

Pro nalezení minimální vzdálenosti je potřeba algoritmem DTW vypočíst matici vzdáleností (cost matrix):

$$D(i, j) = |X[i] - Y[j]| \tag{3.2}$$

Tato matice D pak obsahuje vzdálenosti pro všechny vzorky signálu X a Y. Z těchto vzdáleností nás zajímají ale jen ty cesty, které mají nejmenší vzdálenost. To lze zjistit výpočtem matice L:

$$L(X, Y) = \frac{1}{2T} min[\sum_{k=1}^{K} d(c(k))w(k)] \tag{3.3}$$

kde $w(k)$ je váhový faktor a lze jej vypočíst vztahem:

$$w(k) = [x(k) - x(k-1)] + [y(k) - y(k-1)] \tag{3.4}$$

Implementace této metody se řeší dynamickým programováním. Obdobnou odvozenou metodou z této základní metody je takzvaná FastDTW [3] [4].

3.2 Analýza nezávislých komponent

Metoda analýzy nezávislých komponent(ICA - Independent component analysis) slouží k separaci nezávislých signálu ze smíchaných signálů. Tato metoda je tedy vhodná k redukci šumu z dat[6].

ICA můžeme definovat následovně: Pokud máme dva smíchané signály

$$x_1(t) = a_{11}s_1 + a_{12}s_2$$

$$x_1(t) = a_{11}s_1 + a_{12}s_2$$

Kde s_1 a s_2 jsou dva nezávislé signály a x_1 a x_2 jsou snímané smíchané signály. Následně lze definovat ICA jako

$$X = A.S + N \tag{3.5}$$

Kde X je matice smíchaných signálů, matice A je matice, která provádí smíchání signálů s maticí S a N je vektor náhodného šumu, v praxi se tento vektor šumu převážně zanedbává, můžeme tedy uvažovat jen

$$X = A.S \tag{3.6}$$

Dalším krokem ICA je výpočet či odhad nezávislosti jednotlivých komponent. Určení nejednoznačnosti komponent, například pomocí hustoty pravděpodobnosti, či výpočtem nekorelovanosti[6] [2].

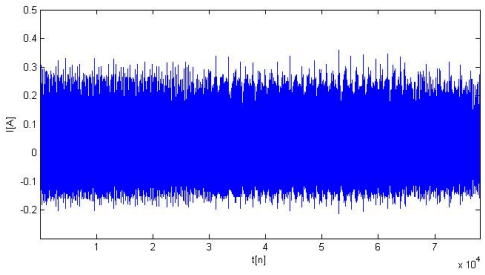
4 VÝSLEDKY STUDENTSKÉ PRÁCE

Náplní této práce je aplikování analýzy hlavních komponent na proudové průběhy algoritmu AES [8] a následovné provedení diferenciální proudové analýzy.

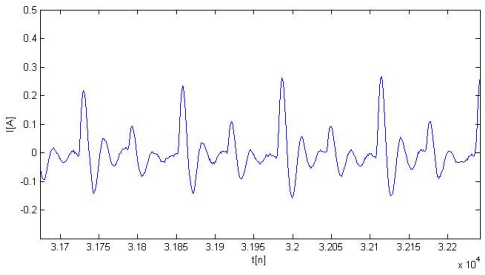
4.1 Interpretace proudových průběhů

V této podkapitole bych chtěl uvést zpracovávaná data z práce [8] a osvětlit problém jejich použití na DPA. Výstup z bakalářské práce, na který jsem navázal je jednoduchá a diferenciální proudová analýza provedena na algoritmu AES implementovaného do třech mikroprocesorů řady ATmega. Tyto analýzy [8] nebyly úspěšné, důvodem je pravděpodobně rozsynchronizování proudových průběhů.

Data set, se kterým budu pracovat, obsahuje 3000 již ořezaných proudových průběhů, vstupní data pro která byly proudové průběhy změřeny a kontrolní matici použitého šifrovacího klíče, data byla změřena s procesorem ATmege16. Při analýze hlavních komponent se pak bude pracovat pouze s maticí proudových průběhů. Matice se vstupními daty je pak zapotřebí pro realizaci DPA [8]. Zde pro představu, graf průběhu pro první proudovou spotřebu

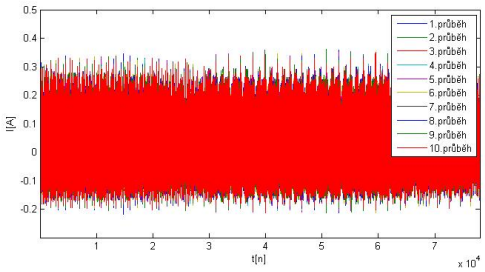


Obr. 4.1: První proudový průběh proudové spotřeby

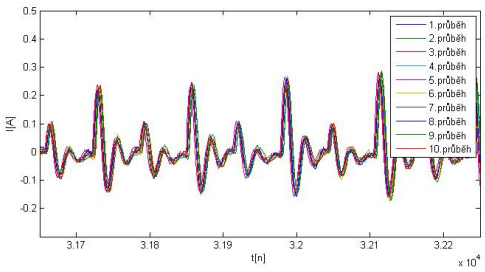


Obr. 4.2: Přiblížený první proudový průběh proudové spotřeby

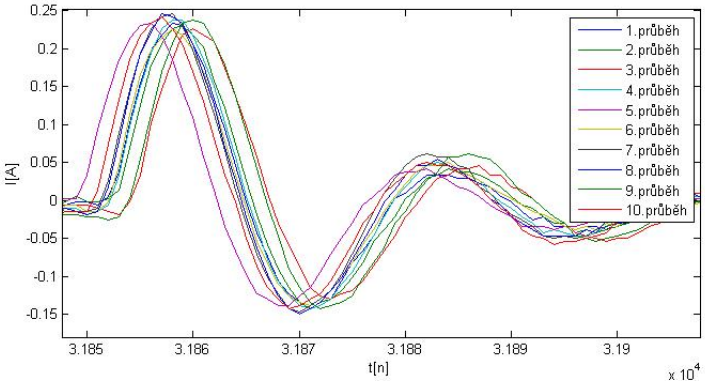
Dále grafy například prvních deseti proudových spotřeb 4.3 4.4 4.5 pro znázornění problému:



Obr. 4.3: Prvních 10 proudových průběhů proudové spotřeby



Obr. 4.4: Přiblížení prvních 10 proudových průběhů proudové spotřeby



Obr. 4.5: Detail prvních 10 proudovýchprůběhů proudové spotřeby

Z obrázku 4.5 je zřejmé, že naměřené proudové průběhy nejsou přesně pod sebou, ale jsou mírně posunuty.

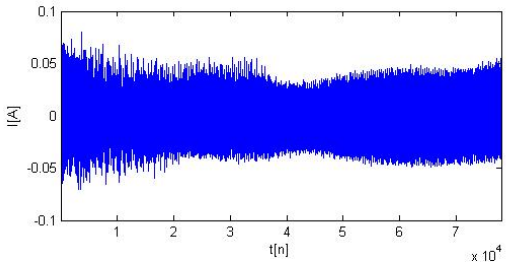
Jak bylo popsáno v práci [8], výstupem DPA je matice průběhů o rozměrech 256x78000 a každý jeden řádek odpovídá průběhu 1 hodnoty bajtu na který je proveden útok, každý bajt může mít hodnotu v rozsahu 0-255. Z definice DPA metodou korelačních koeficientů vyplývá, že průběh pro hodnotu bajtu odpovídající hodnotě tajného klíče se bude od ostatních průběhů lišit [8].

Pro představu zde uvedu útok na druhý a například 15. bajt tajného klíče.

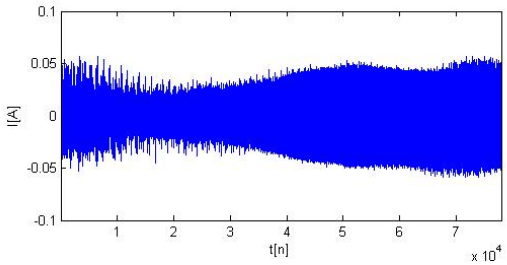
Hodnota šifrovacího klíče pro druhý bajt je 25, měl by tedy dvacátý šestý proudový průběh vyčnívat, 26 protože první průběh je pro hodnotu klíče 0. Průběh klíče pro hodnotu 25 je zobrazen na obrázku 4.9, při porovnání s průběhy klíče pro hodnoty 22, 23 a 24 zobrazených na obrázcích 4.6 4.7 a 4.8, jde vidět, že se tento 26 průběh nijak výrazně neliší, DPA je tedy neprůkazná [8].

Pro ověření jsem provedl diferenciální proudovou analýzu i pro 15. bajt klíče. Následující řada obrázků zobrazuje proudové průběhy pro hodnoty 15. bajtu šifrovacího klíče.

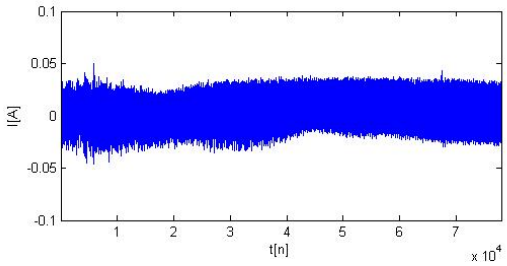
Obrázek 4.10 ukazuje průběh pro hodnotu 128, obrázek 4.11 náleží hodnotě 129,



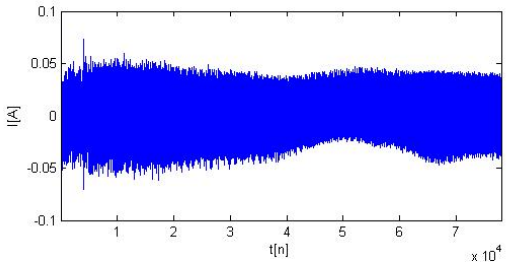
Obr. 4.6: Průběh pro 2. bajt klíče s hodnotou 23



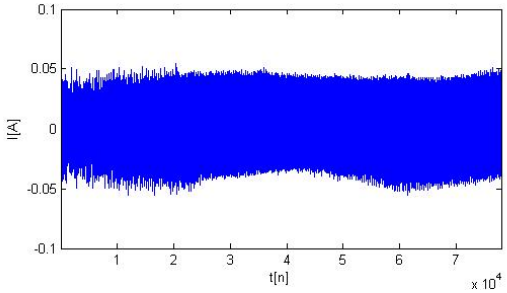
Obr. 4.7: Průběh pro 2. bajt klíče s hodnotou 24



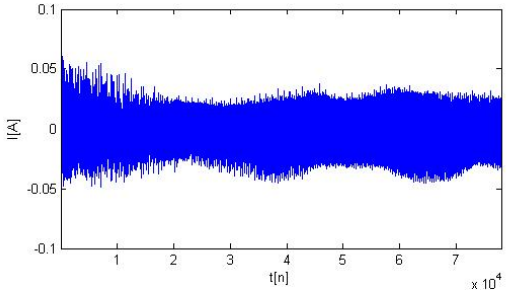
Obr. 4.8: Průběh pro 2. bajt klíče s hodnotou 25



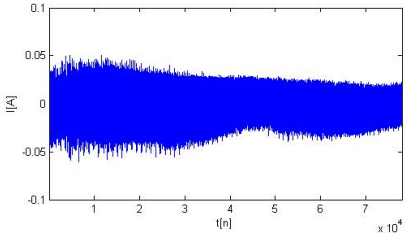
Obr. 4.9: Průběh pro 2. bajt klíče s hodnotou 26



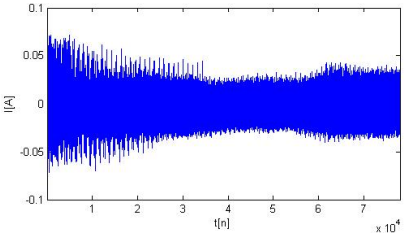
Obr. 4.10: Průběh pro 15. bajt klíče s hodnotou 129



Obr. 4.11: Průběh pro 15. bajt klíče s hodnotou 130



Obr. 4.12: Průběh pro 15. bajt klíče s hodnotou 131



Obr. 4.13: Průběh pro 15. bajt klíče s hodnotou 132

obrázek 4.12 pro hodnotu 130 a obrázek 4.13 odpovídá hodnotě klíče 131. Hodnota klíče pro 15. bajt je 130, takže by měl vyčnívat průběh na obrázku 4.12, ale není tomu tak.

4.2 Realizace analýzy hlavních komponent

Jak bylo ukázáno v předcházející kapitole, na změřené proudové průběhy nelze aplikovat proudovou analýzu. Důvodem je pravděpodobně špatně synchronizovaný signál, tzn. jednotlivé změřené průběhy jsou proti sobě mírně posunuty. Tento problém lze řešit předzpracováním dat před proudovou analýzou. V této části je provedena analýza hlavních komponent (PCA) na proudové průběhy za účelem odstranění šumu, redukcí zašumělých dimenzí a zvýraznění užitečných informací.

Výpočet hlavních komponent lze v matlabu provést několika způsoby, například existuje hned několik příkazů pro výpočet pca, jsou to příkazy:

$$[COEFF, SCORE, latent, tsquared, explained, mu] = pca(X)$$

$$[COEFF, SCORE, latent, tsquare] = princomp(X)$$

$$[COEFF, latent, explained] = pcacov(cov(X))$$

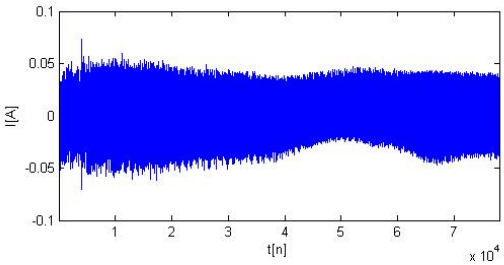
kde $COEFF$ představuje matici vlastních vektorů seřazených sestupně podle důležitosti tj. podle velikosti vlastních čísel, matice $SCORE$ obsahuje data ze vstupní matice X reprezentované hlavními komponentami na místo původních os, tato matice má rozměry X^T. Vektor *latent* obsahuje vlastní čísla náležící vlastním vektorům matice $SCORE$, vlastní čísla jsou tedy seřazeny od největšího po nejmenší. Matice *tsquared* vrací T^2 hotelingovou statistiku pro každou dimenzi matice X. Výstup *explained* obsahuje procentuální vyjádření totálního rozptylu vyčerpaného každou hlavní komponentou. A výstup *mu* obsahuje odhadovaný průměr každé proměnné v X. Výstupy *explained* a *mu* funkce *pca* nelze použít na starších verzích Matlabu.

Podle výstupu funkcí můžeme vidět, že se od sebe liší jen přetíženými výstupy.

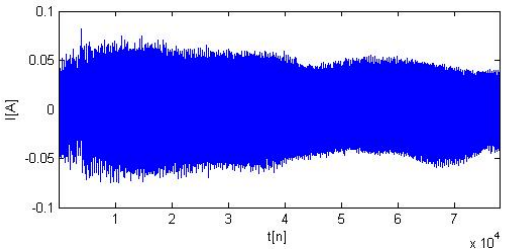
Další užívanou praktikou výpočtu PCA je možnost vypočítat neseřazené vlastní vektory a čísla příkazem $[eigenvectors, eigenvalues] = eig(X)$, kde $eigenvectors$ je matice vlastních vektorů matice X, a $eigenvalues$ je matice obsahující neseřazené vlastní čísla náležící vlastním vektorům matice $eigenvectors$. Nevýhodou tohoto příkazu je nutnost seřazení vlastních vektorů(v novějších verzích programu jsou vektory seřazeny vzestupně). Hodnoty vlastních vektorů z funkcí *princomp* a *eig* se liší ve znaménkách vlastních vektorů. U této medy je díle nastává nutnost postupného výpočtu hlavních komponent dle matematické teorie z 2. Rozhodl jsem se tedy zvolit si tuto modu výpočtu pro svoji práci. [11] [10] [12]

4.2.1 Zpracovávaná data

Dále jsem se rozhodl provádět pca pouze na 2000 průbězích z celkově naměřených 3000, důvodů je několik. Podle DPA je sice útok teoreticky úspěšnější v závislosti na počtu změřených průběhů, ale v praxi jde úspěšný útok provést již od přibližně 200 až 500 průběhů. Právě proto by bylo zbytečné a náročné na výpočetní čas a výkon pracovat s tak objemnými daty. Pro potvrzení tohoto tvrzení, zde uvedu příklad provedeného DPA na průběhu pro původní data a data zmenšená na 2000 průběhů.



Obr. 4.14: Průběh pro 2. bajt klíče s hodnotou 26 pro 3000 průběhů



Obr. 4.15: Průběh pro 2. bajt klíče s hodnotou 26 pro 2000 průběhů

Z porovnání obrázků 4.14 a 4.15 je zřejmé, že odebraná data měla na DPA útok minimální vliv, tvar průběhů je téměř totožný. Dále jsem se z důvodu přehlednosti a rychlosti rozhodl následující PCA prezentovat na průbězích seřízlích jen na tu část šifrovacího cyklu kdy dochází k přičtení klíče jak vyplývá z teorie DPA. Prezentované výsledky jsou tedy provedeny na částech průběhu od bodu 28000 do bodu 66000, tedy s daty o rozměrech 2000x38001. Stejnou analýzu jsem provedl i s celými daty s obdobnými výsledky.

4.2.2 Výstup PCA

Při provedení PCA na 2000 proudových průběhů dostaneme 2000 hlavních komponent (Principal component – PC) seřazených podle důležitosti. Jen prvních několik komponent má největší vypovídací schopnost o datech a zbylé komponenty zachycují převážně šum. Po výběru hlavních komponent, které si ponecháme, máme 2 možnosti jak naložit s komponentami. Můžeme buďto provést jednoduchou transformaci dat a tím teoreticky obsáhnout vypovídací schopnost datového souboru do méně dimenzí, tato metoda není vhodná pro proudovou analýzu. Nebo můžeme provést redukci dimenzí po porovnání jednotlivých hlavních komponent s jednotlivými průběhy, tento postup je objasněn a proveden níže.

Existuje několik kritérií, podle kterých můžeme vybrat hlavní komponenty.

- 1. ponecháme si pouze ty komponenty, které mají hodnotu vlastního čísla větší než 1.
- 2. Grafickou metodou, vyčtením ze sutinového grafu (scree plot), viz níže.
- 3. Můžeme vybrat pouze ty komponenty jejichž variabilita přesahuje určitou mez.

Tabulka 4.1 ukazuje prvních 10 komponent.

Tab. 4.1: Popis prvních 10. komponent

Hlavní komponenta	**Velikost vlastního čísla**	**Vyčerpání rozptylu[%]**
1.	3,39439	29,79373
2.	2,57475	22,59951
3.	1,76710	15,51045
4.	1,18892	10,43553
5.	0,85189	7,47738
6.	0,59256	5,20113
7.	0,27765	2,43700
8.	0,24195	2,12367
9.	0,12371	1,08584
10.	0,08071	0,70839
Celkem rozptylu		97,37263

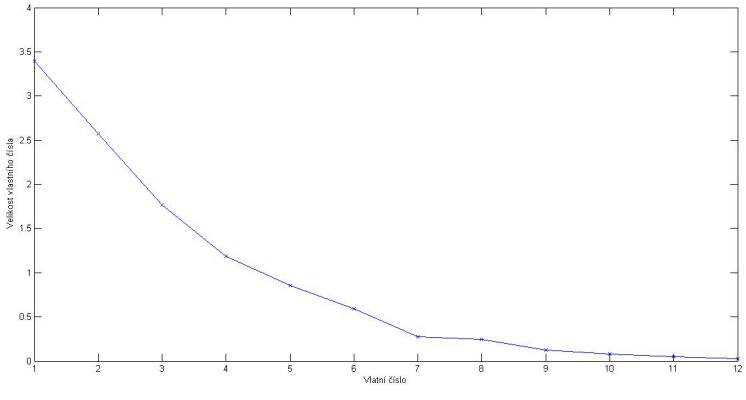
Těchto prvních 10 komponent zabírá 97,37 % celkového rozptylu dat, mají tedy teoreticky největší vypovídací vlastnost o celém datovém souboru. Prvních 6 komponent obsahuje přibližně 91 % rozptylu, zbylé komponenty pravděpodobně zachycují převážně šum a obsahují jen velmi malé množství informace.

Při dodržení 1. kritéria bychom tedy vybraly první 4 komponenty, a při dodržení 3. kritéria 6 komponent.

Ze sutinového grafu na obrázku 4.16 vybereme pouze komponenty od jedné po komponentu u které se vlastní číslo v grafu přestává lomit a přechází do pozvolného klesání. Graf se láme v pátém až šestém bodě, takže nám s tímto kritériem stačí vybrat pouze prvních 5 hlavních komponent. Dále jsem tedy pracoval se čtyřmi, pěti i šesti komponentami.

Výběr dat a redukce dimenze

Problém nastává při prezentaci daných komponent a určení, jaká data daná komponenta zastupuje. Toto určení respektive porovnání můžeme provést buďto graficky



Obr. 4.16: Sutinový graf prvních pro prvních 12 komponent

srovnáním dané komponenty s jednotlivými průběhy nebo můžeme provést korelaci vybraných komponent s průběhy.

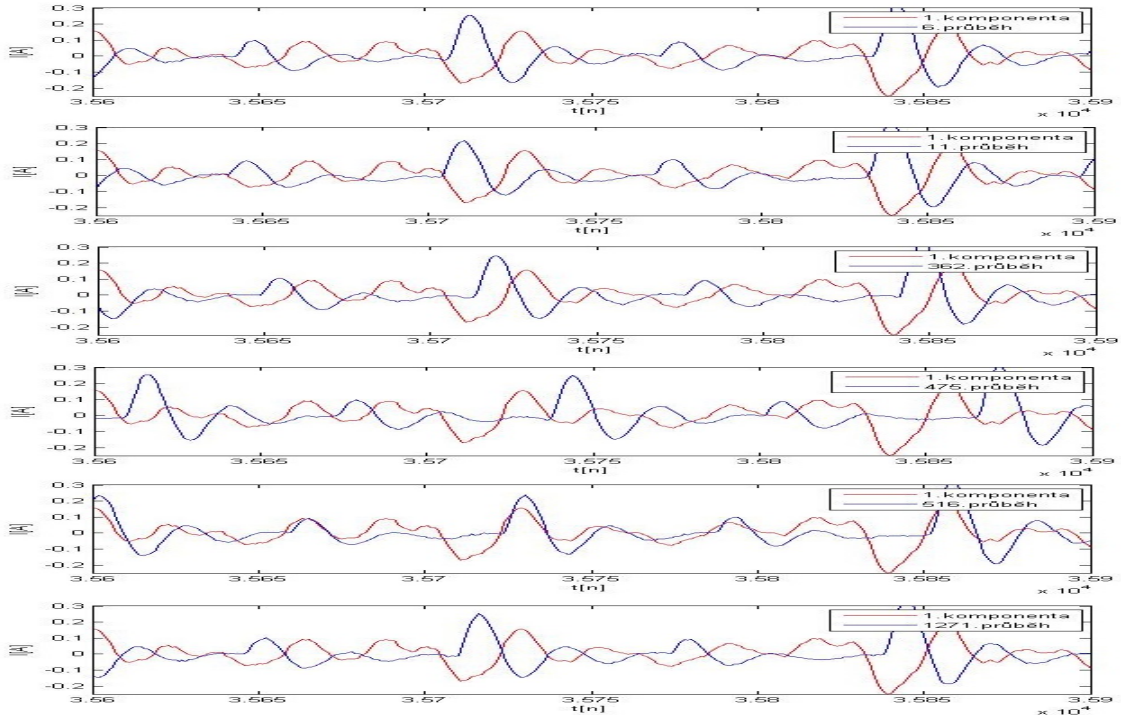
Dále jsem tedy provedl korelaci prvních 5 komponent s průběhy. Jak již bylo řečeno výše výstupem korelace je hodnota od -1 do 1 a jak bylo vysvětleno v 2.1 na kovarianci. Zajímají nás tedy ty výsledky korelací mezi průběhy a jednotlivými komponentami, které se v absolutní hodnotě blíží k 1. Provedl jsem tedy korelace a testoval s výběrem průběhů pro danou komponentu, které korelovaly v rozmezí od 0,65 do 0,88 v absolutní hodnotě. Skripty použité na korelaci dat s komponentou, následný výběr dat a konečnou redukci dimenze jsou součástí přiloženého DVD.

Tab. 4.2: Přehled průběhů zachycených hl. komponentami pro mez korelace

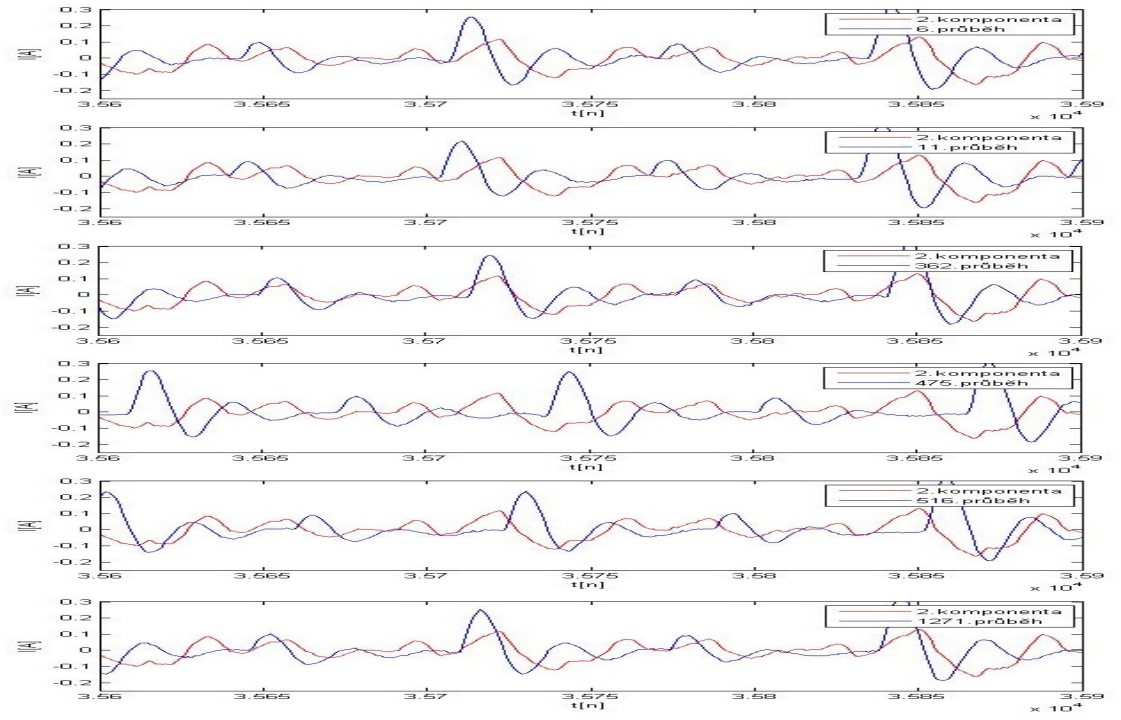
PC	**k>0.65**	**k>0.7**	**k>0.75**	**k>0.8**	**k>0.85**	**k>0.88**
1.	276	242	205	170	115	63
2.	446	281	123	-	-	-
3.	271	156	67	-	-	-
4.	21	-	-	-	-	-
5.	117	-	-	-	-	-
Celkem	1319	679	395	170	115	63

V tabulce 4.2 je zaznamenáno kolik průběhů můžeme přiřadit dané komponentě při požadované úrovni korelace. Pro názornost uvedu ukázky vybraných šesti průběhů porovnaných s prvními třemi komponentami na obrázcích 4.17, 4.18 a 4.19.

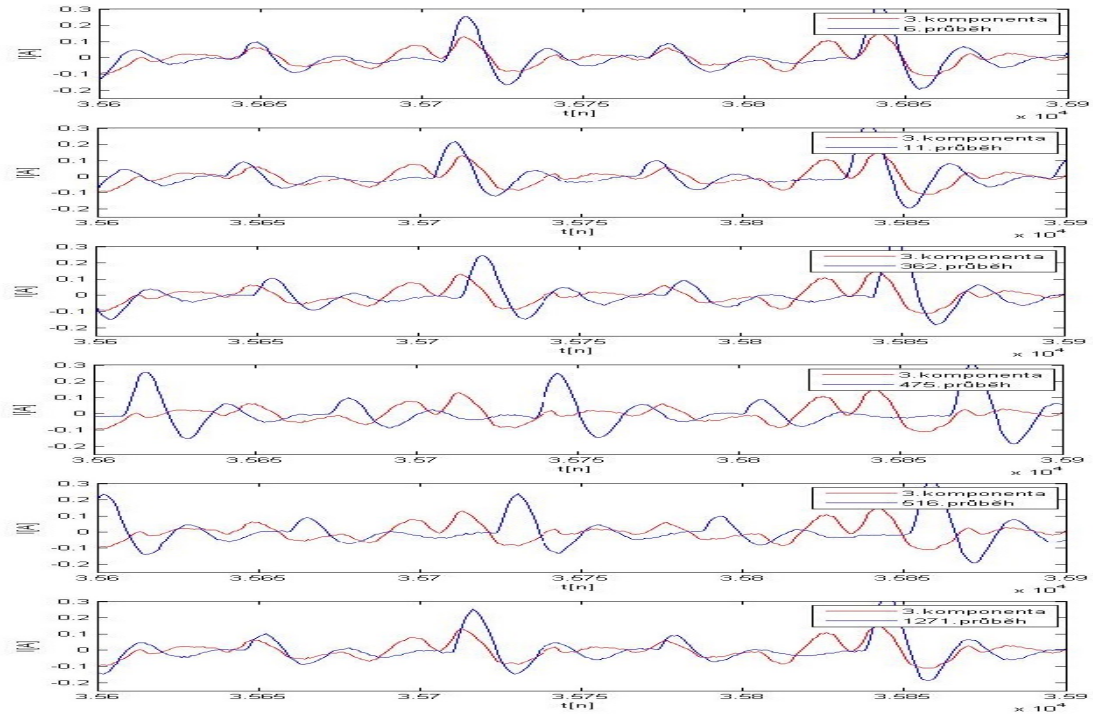
Hodnoty korelací těchto průběhů s jednotlivými komponentami jsou uvedeny na v tabulce 4.3.



Obr. 4.17: Porovnání 1. komponenty s průběhy

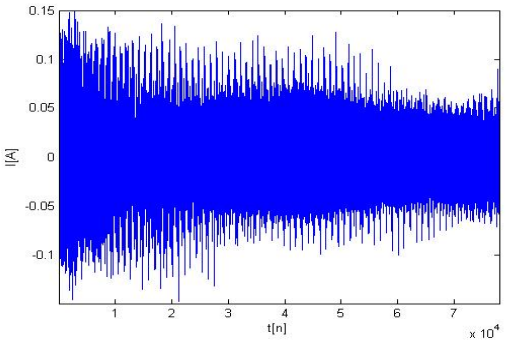


Obr. 4.18: Porovnání 2. komponenty s průběhy

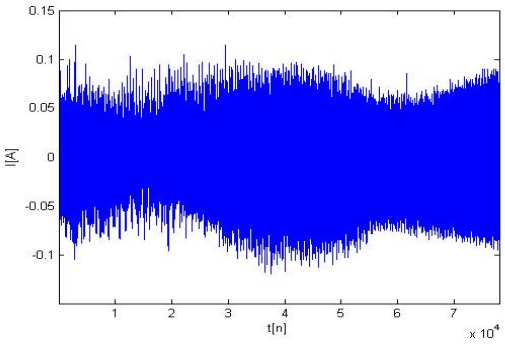


Obr. 4.19: Porovnání 3. komponenty s průběhy

Pro všech 5 mezí korelace jsem extrahoval nové datové soubory, průběhy zachycené jednotlivými komponentami bylo nutno vůči sobě zarovnat, protože každá komponenta zachycuje převážně data posunuté od dat zachycené jinou komponentou. Dále je také vhodné odstranit ty průběhy, které zachycuje více komponent a přiřadit je jen jedné podle vhodných kritérií.



Obr. 4.20: Průběh pro 1. bajt klíče s hodnotou 10



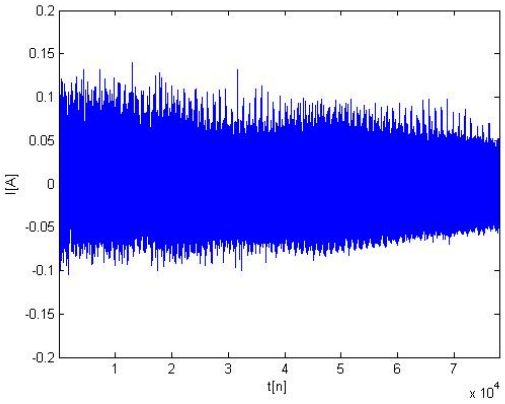
Obr. 4.21: Průběh pro 1. bajt klíče s hodnotou 11

Následně jsem s takto redukovanými daty provedl proudovou analýzu. Výsledky ani jednoho DPA útoku nebyly úspěšné. Pro úplnost zde uvedu výsledky DPA pro

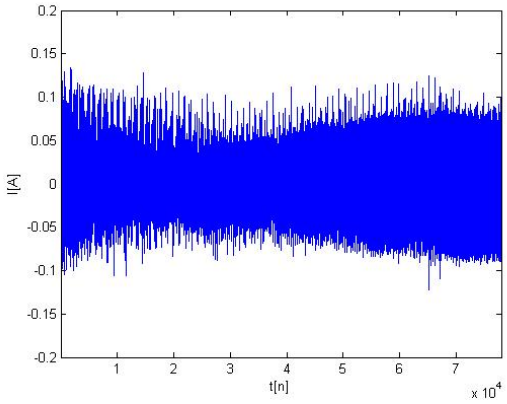
Tab. 4.3: Přehled průběhů zachycených hl. komponentami pro mez korelace

Číslo průběhu	PC1	PC2	PC3
6	-0,8856	0,2978	0,7824
11	-0,6151	-0,7558	0,1074
362	-0,5042	0,6647	0,2594
475	0,0373	-0,769	-0,1717
516	0,4632	0,0883	-0,6151
1271	-0,8515	0,5149	0,6702

datový soubor s ponechanými průběhy korelujícími s komponentami nad 0,65 v absolutní hodnotě na obrázcích 4.20 a 4.21 pro 1.bajt klíče a na obrázcích 4.22, 4.23 pro 16. bajt klíče. Hodnota tajného klíče pro 1. bajt je 10 což odpovídá průběhu 4.21, který se nijak významně neodlišuje od 4.20. Obdobně hodnota tajného klíče pro 16. bajt je 179 což odpovídá průběhu 4.23, který také nijak nevyčnívá.



Obr. 4.22: Průběh pro 16. bajt klíče s hodnotou 179

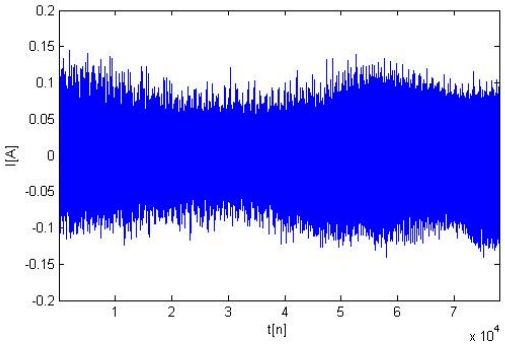


Obr. 4.23: Průběh pro 16. bajt klíče s hodnotou 180

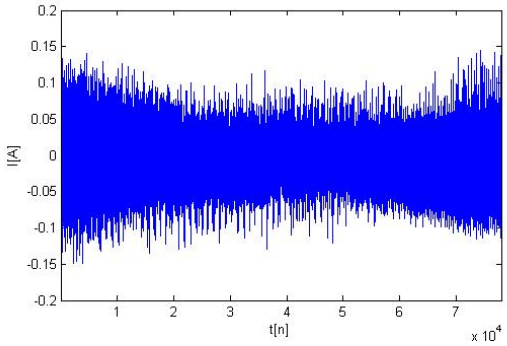
A výsledek DPA na míru korelace 0,75 v absolutní hodnotě pro 16. bajt klíče 4.25, 4.25.

Pro úplnost zde uvedu ukázky průběhů 4.26, 4.27, 4.28.

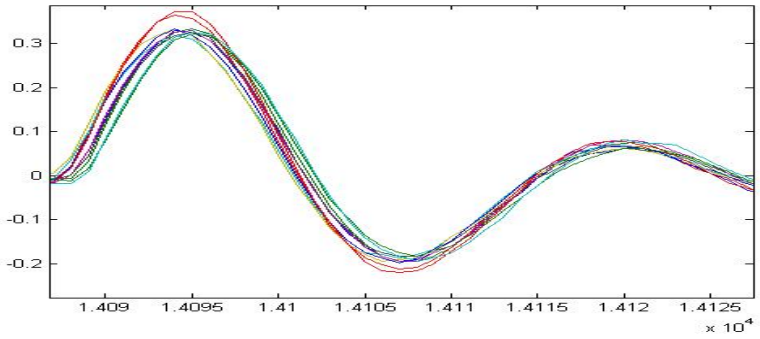
Z grafu 4.26 můžeme vidět, že proudové průběhy zachycené hlavní komponentou jsou pod sebou tak jak potřebujeme pro proudovou analýzu. Zatímco ostatní komponenty, které s daty korelují méně se pod sebou opět rozcházejí. Aby mohla být proudová analýza úspěšná, potřebovali bychom pracovat s proudovými průběhy korelujícími minimálně z 80 %, což odpovídá pouze 170 proudovým průběhům zachycených hlavní komponentou. Z teorie o proudové analýze z [7] vyplývá, že pro úspěšný útok pomocí diferenciální proudové analýzy potřebujeme minimálně 200 až



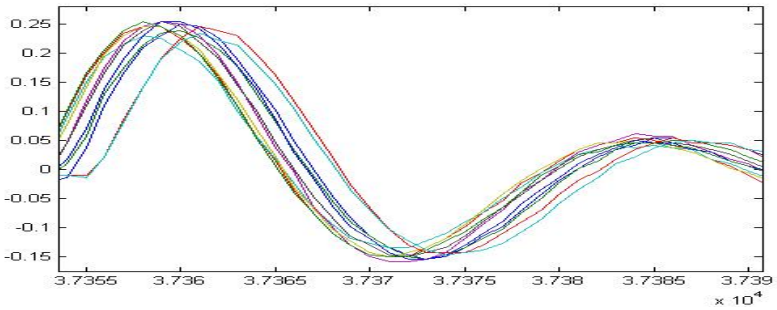
Obr. 4.24: Průběh pro 16. bajt klíče s hodnotou 179



Obr. 4.25: Průběh pro 16. bajt klíče s hodnotou 180



Obr. 4.26: 10 průběhů vybraných hlavní komponentou při vysoké korelaci

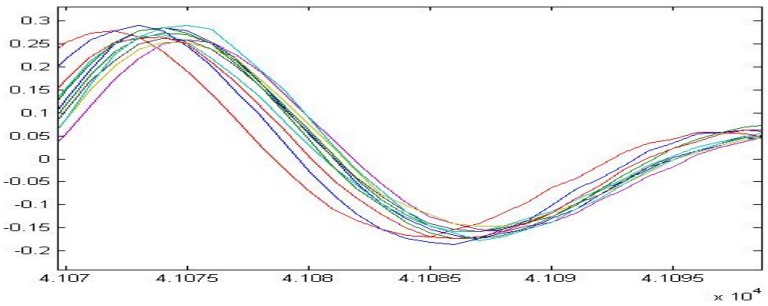


Obr. 4.27: 10 průběhů vybraných druhou komponentou při korelaci do 0.75

500 nezašumělých proudových průběhu.

Možné důvody neúspěchu

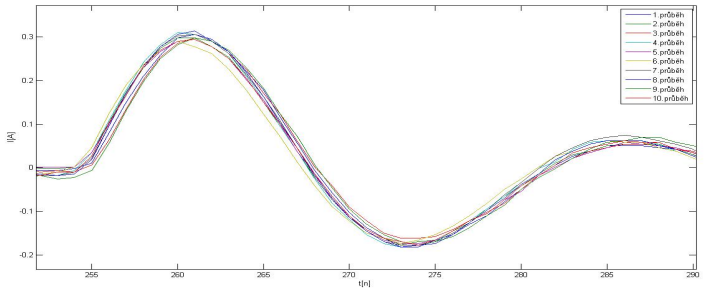
Možnou příčinou mohou být nerovnoměrně rozsynchronizované průběhy na časové ose. Na obrázcích 4.29, 4.30, 4.31 jsou zobrazeny původní proudové průběhy v de-



Obr. 4.28: 10 průběhů vybraných druhou třetí při korelaci do 0.65

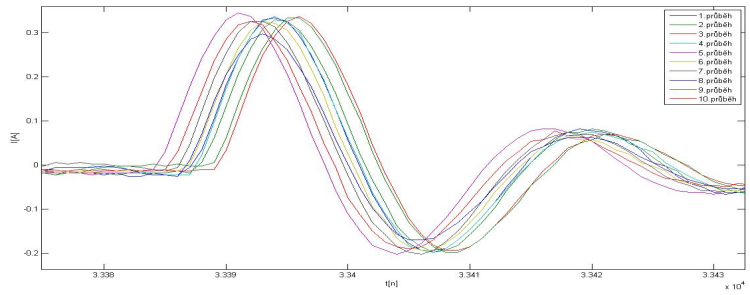
tailu na začátku, uprostřed a na konci průběhu šifrování. Důvodem může být také pravděpodobně zvolená špatná metoda předzpracování dat PCA na tento datový soubor.

Lze tedy dojít k závěry, že proudové průběhy nejsou nezarovnány konstantní délkou, ale tato délka mezi špičkami se během průběhu šifrování různě mění. Z toho vyplývá, že u takto nesynchronizovaných dat nelze úspěšně aplikovat PCA, protože počet podobných, vysoce korelovatelných proudových průběhů je příliš malý na úspěšnou diferenciální proudovou analýzu.

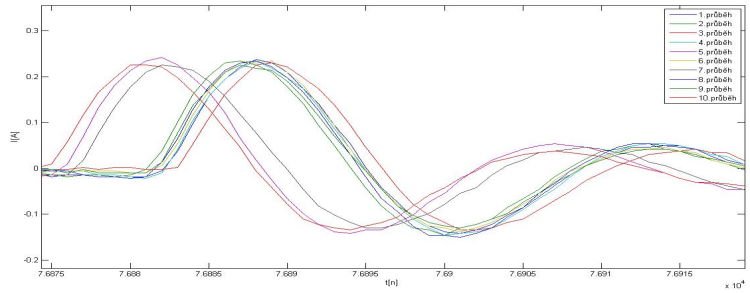


Obr. 4.29: Prvních 10 proudových průběhů na začátku šifrování

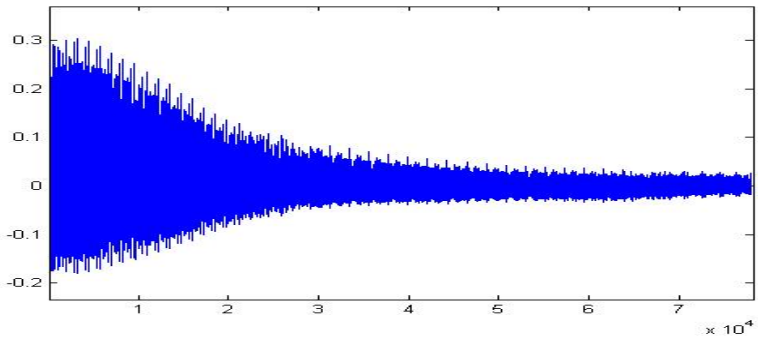
Na takto rozsynchronizovaná data by bylo vhodnější použít metodu borcení časové osy (DTW-Dynamic time warping) viz 3.1. Také dalším možným problémem s daty může být střední hodnota ve většině průběhu přibližující se k nule. Na obrázku 4.32 je zobrazen průměr ze všech 2000 průběhů. Amplituda by měla být pohledově přibližně konstantní, ale není tomu tak, to může působit problémy při proudové analýze, může to být příčina její neúspěšnosti.



Obr. 4.30: Prvních 10 proudových průběhů uprostřed šifrování



Obr. 4.31: Prvních 10 proudových průběhů na konci šifrování



Obr. 4.32: Průměr hodnot že všech průběhů

5 ZÁVĚR

Cílem této práce byla realizace analýzy hlavních komponent na změřených proudových průbězích kryptografického modulu algoritmu AES. V teoretické části je úvod do problematiky kryptoanalýzy proudovým postranním kanálem a kryptoanalýzy obecně. Dále je pak vypracován teoretický postup analýzy hlavních komponent a postup realizace PCA v Matlabu.

V praktické části je pak provedena analýza hlavních komponent v prostředí Matlab na obdržené průběhy a následná aplikace DPA na takto data redukovaná pomocí hlavních komponent. Tento útok se nezdařil a nepovedlo se získat hodnotu tajného klíče. Důvodem může být pravděpodobně zvolená špatná metoda předzpracování dat PCA na tento datový soubor. Na průběhy by bylo pravděpodobně vhodnější aplikovat metodu borcení časové osy DWT. Použité průběhy a skripty jsou součástí přiloženého DVD.

LITERATURA

[1] BATINA, Lejla, Jip HOGENBOOM a Jasper G.J. VAN WOUDENBERG. *Getting More from PCA: First Results of Using Principal Component Analysis for Extensive Power Analysis.* [online]. [cit. 2014-12-15]. Dostupné z: https://www.riscure.com/benzine/documents/ctrsa12.pdf

[2] BOHY, Lilian, Michael NEVE, David SAMYDE a Jean-Jacques QUISQUATER. *Principal and Independent Component Analysis for Crypto-systems with Hardware Unmasked Units* [online]. Louvain, 2003 [cit. 2015-06-01]. Dostupné z: http://www.researchgate.net/profile/Jean-Jacques_Quisquater/publication/228997254_Principal_and_Independent_Component_Analysis_for_Crypto-systems_with_Hardware_Unmasked_Units/links/09e4150b9ce99ca357000000.pdf.

[3] VAN WOUDENBERG, Jasper G. J., Marc F. WITTEMAN a Bram BAKKER. Improving Differential Power Analysis by Elastic Alignment. In: VAN WOUDENBERG, Jasper G. J., Marc F. WITTEMAN a Bram BAKKER. *Improving Differential Power Analysis by Elastic Alignment* [online]. 2011 [cit. 2015-06-01]. Dostupné z: https://www.riscure.com/documents/improving_differential_power_analysis_by_elastic_alignment.pdf?1378979585

[4] KEOGH, Eamonn J. a Michael J. PAZZANI. Scaling up Dynamic Time Warping to Massive Datasets. In: *Scaling up Dynamic Time Warping to Massive Datasets* [online]. 1999 [cit. 2015-06-01]. Dostupné z: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.81.7594&rep=rep1&type=pdf

[5] Kocher, P. C.; Jaffe, J.; Jun, B.: Differential Power Analysis. In CRYPTO '99: Proceedings of the 19th Annual International Cryptology Conference onAdvances in Cryptology, London, UK: Springer-Verlag, 1999, ISBN 3-540-66347-9, s. 388–397.

[6] KONOPKA, Ondřej. *Analýza nezávislých komponent* [online]. Praha, 2004 [cit. 2015-06-02]. Dostupné z: http://amber.feld.cvut.cz/bio/konopka/file/LBR-semestralka.pdf. Semestrální projekt. ČVUT.

[7] Mangard, S.; Oswald, E.; Popp, T.: Power Analysis Attacks: Revealing the Secrets of Smart Cards (Advances in Information Security). Secaucus, NJ, USA:Springer-Verlag New York, Inc., 2007, ISBN 0387308571.

[8] MAREK, Pavel *Realizace diferenciální proudové analýzy*: bakalářská práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních

technologií, Ústav telekomunikací, 2014. 67 s. Vedoucí práce byl Ing. Zdeněk Martinásek, PhD.

[9] MARTINÁSEK, Zdeněk. *Kryptoanalýza postranními kanály*: dizertační práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2013. 129 s. Vedoucí práce byl doc. Ing. Václav Zeman, Ph.D.

[10] Mathworks help: pca. [online]. [cit. 2014-12-15]. Dostupné z: `http://www.mathworks.com/help/stats/pca.html`

[11] Mathworks help: pcacomp. [online]. [cit. 2014-12-15]. Dostupné z: `http://www.mathworks.com/help/stats/princomp.html`

[12] Mathworks help: pcacov. [online]. [cit. 2014-12-16]. Dostupné z: `http://www.mathworks.com/help/stats/pcacov.html`

[13] Smith, L.I.: *A tutorial on principal components analysis* (February 2002), Dostupné z: `http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf`

[14] WALCZYSKO,Martin. *Analýza EEG signálu pomocí analýzy hlavních komponent (PCA)*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2008. 44 s. Vedoucí bakalářské práce Ing. Milan Rychtárik

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

AES	Advancet encription standard – Pokročilý šifrovací standard
Cov	Covariance
CMOS	Complementary Metal Oxide Semiconductor
DPA	Differencial power analysis – Diferenciální proudový analýza
ICA	Independ component analysis – Analýza nezávislích komponent
PCA	Principal component analysis – Analýza hlavních komponent
SPA	Simple power analysis – Jednoduchá proudová analýza
DTW	Dinamic time warping – Borcení časové osy