Crypto-World

Informační sešit GCUCMP

ISSN 1801-2140

Ročník 10, číslo 6/2008 **15. červen 2008**

6/2008

Připravil: Mgr. Pavel Vondruška

Sešit je přednostně distribuován registrovaným čtenářům.
Starší sešity jsou dostupné na adrese
http://crypto-world.info
(1240 registrovaných odběratelů)

Obsah:

A. RFID: Co to vlastně máme v kapse?

Martin Hlaváč (MFF UK, PPF Banka, a.s., hlavm1am@artax.karlin.mff.cuni.cz),
Tomáš Rosa (eBanka, a.s. trosa@ebanka.cz)

Abstrakt

Napadlo vás někdy, jak je asi složité vyrobit duplikát k vašemu modernímu bezkontaktnímu klíči od kanceláře, garáže či domu? Líbí se vám pomyšlení, že možná stačí letmý kontakt s útočníkem ve výtahu či MHD a už má vaše klíče? Příspěvek přístupnou formou vysvětluje základní fyzikální a informatické principy RFID. Kromě slabin obecného rázu ukazuje i na zranitelnosti spojené s nějakou konkrétní, hojně rozšířenou technologií, jako jsou čipy přístupových systémů v pásmu LF, platforma MIFARE, atp. Na příkladu elektronických cestovních pasů si představíme přepojovací útok, který i přes nespornou přímočarost a jednoduchost základní ideji představuje jednu z největších hrozeb moderním aplikacím RFID. Pozornost bude věnována též nastupující generaci duálních čipových karet a fenoménu NFC.

Klíčová slova: bezkontaktní, čipy, NFC, RFID, útoky

1. Úvod

Zkratka RFID (Radio Frequency IDentification) označující radiofrekvenční identifikaci subjektů je dodnes často považována za věc budoucnosti. Metody tohoto druhu jsou přitom známy přinejmenším od 2. světové války, kde byly vyvíjeny a nasazovány ruku v ruce se zařízením zvaným RADAR (RAdio Detection And Ranging), které asi není nutné nějak zvlášť představovat. RFID v té době mělo podobu speciálních odpovídačů, jimiž byla vybavena spřátelená letadla té které strany. Tato zařízení většinou na pokyn pilota odeslala určitý rádiový signál, který domovský RADAR zachytil a zobrazil například jako specifický stín vedle bodu znázorňujícího detekované letadlo. Tím byl stroj pro obsluhu velínu identifikován jako spřátelený.

Podobně jako RADAR i RFID prošlo od 2. světové války dlouhým vývojem. Možná by nás překvapilo, co všechno vlastně splňuje definici tohoto zařízení. Namátkou vyberme třeba mobilní telefon GSM, s jehož pomocí se držitel mimo jiné prostřednictvím rádiového přenosu identifikuje v síti operátora. Do kategorie RFID dále patří zařízení pro podporu mýtných systémů montovaná na palubní desky nákladních automobilů. To jsme ovšem stále ve světě „velkých“ zařízení. Naše pozornost bude v tomto příspěvku zaměřena na zcela specifickou podmnožinu označovanou jako *pasivní* RFID. Výraz pasivní zde znamená, že příslušné moduly nejsou vybaveny autonomními zdroji elektrické energie. Tu totiž získávají přímo z pole terminálu, se kterým komunikují. Na první pohled se zdá, že tímto způsobem lze udržet v chodu jen velmi jednoduché čipy, které snad ani nemohou poskytovat bezpečí a komfort, na který jsme zvyklí například u klasických (kontaktních) čipových karet. Nicméně technologie výroby integrovaných obvodů je dnes již tak pokročilá, že zrovna moderní bezkontaktní karty se mohou co do schopností směle měřit s těmi kontaktními. Příkladem budiž aplikace elektronického cestovního pasu [31].

Skutečnost, že na trhu existují kvalitní čipy pro RFID, ovšem ještě zdaleka neznamená, že je konstruktéři vždy použijí, případně že je použijí správně. S analogickými problémy se ostatně často potýkáme i v jiných oblastech bezpečnosti. Snahou tohoto příspěvku je nenáročnou formou představit oblast pasivního RFID a na základě praktických zkušeností upozornit na potenciální rizika spojená s jeho nasazením a provozem. Následující členění je takovéto: Ve

druhé části provedeme klasifikaci čipů podle rádiových pásem, ve kterých pracují. Nadále se přitom zaměříme na čipy pásem LF (Low Frequency) a HF (High Frequency), neboť u nich můžeme s výhodou využít řady podobných rysů. Pasivní čipy pásma UHF (Ultra High Freqeuncy) bohužel překračují rámec tohoto příspěvku, nicméně jejich základní charakteristiky v části dva uvedeme. Ve třetí části navážeme jednoduchým seznámením s elementárními radioelektronickými principy, na kterých jsou čipy LF a HF založeny. Všimneme si přitom zejména, jak situace vypadá z pohledu útočníka. Části čtyři, pět a šest jsou postupně věnovány charakteristickým zástupcům čipů, se kterými se dnes můžeme v praxi obvykle potkat. Cílem je demonstrovat na nich běžné hrozby, které ne vždy bývají korektně ošetřeny. Sedmá část v krátkosti představuje hlavní trendy této oblasti, kterými jsou duální čipová karta, bezkontaktní platební karta a rozhraní NFC. Naše seznámení s pasivním RFID uzavřeme v části osm.

2. Radio-klasifikace transpondérů

V souladu s běžnou terminologií budeme nadále považovat výrazy čip RFID a transpondér RFID za synonyma. Rovnou na tomto místě je vhodné poznamenat, že pro udržení souladu s terminologií studované oblasti budeme poněkud v rozporu s přístupem v oblasti informační bezpečnosti používat výraz *identifikace* jak pro rozpoznání totožnosti subjektu tak i pro její bezpečné prokázání (čili autentizaci [21]). To, je-li identifikace doprovázena ještě kvalitní autentizací, budeme odlišovat vyjádřením její síly. Za slabou identifikaci budeme považovat tu, která je založena jen na prostém oznámení identity subjektu bez nějakého dalšího bezpečného prokazování. Za bezpečnostní opatření přitom v souladu s běžnou praxí nepovažujeme to, zda je oznámení prováděno prostřednictvím protokolu s neveřejným popisem, atp. Středně silná identifikace je taková, která je sice doplněna nějakou autentizační metodou, avšak její síla je nepříliš přesvědčivá. Typickým příkladem jsou karty MIFARE používající neveřejný algoritmus s délkou klíče 48 bitů. Za silnou identifikaci budeme považovat tu, která je doplněna přesvědčivou autentizací založenou na respektovaných kryptografických standardech. Příkladem může být metoda použitá k identifikaci čipu elektronického pasu, která se označuje jako aktivní autentizace. V tabulce níže označujeme jednotlivé úrovně zažitými symboly L, M, H pro nízkou, střední a silnou identifikaci.

Ačkoliv čistě teoreticky tomu tak být nemusí, ukazuje se, že vhodným kriteriem pro rozdělení čipů pasivního RFID pro účely analýzy bezpečnosti je jejich pracovní frekvence. Důvody jsou patrně historické, kdy první čipy pracovaly v pásmu LF a byly to v podstatě jen sériové paměti. Současně s tím, jak začali konstruktéři vylepšovat bezpečnostní vlastnosti čipů, docházelo zároveň k postupnému přechodu na výhodnější i když technicky náročnější pásmo HF. To se postupně stalo synonymem pro funkčně i bezpečnostně vyzrálejší modely. Nejvyšší míru bezpečnosti můžeme obecně očekávat u čipů z kategorie *karty s vazbou na blízko*, o které se více zmíníme v části 3. Poněkud paralelně k tomu všemu probíhal vývoj čipů pro pásmo UHF, jejichž prvořadým účelem bylo a je nahradit mechanické štítky s čárovými kódy něčím, co lze jednak číst na větší vzdálenost (řádově metry), jednak lépe chránit před nepříznivými vlivy prostředí. Bezpečnost byla vždy jaksi nejvýš druhá v pořadí a její úroveň je v podstatě stejná jako u čipů pásma LF.

Zmiňme se ještě v krátkosti o provedení jednotlivých čipů. Zde lze říci, že typ provedení a typ čipu spolu prakticky téměř nesouvisí. Výrobci jsou dnes obvykle schopni dodat příslušný čip coby plastovou kartu, samolepku, lístek na stadion, přívěsek ke klíčům, kožní implantát, atp. Platí známé klišé, že fantazii se meze rozhodně nekladou.

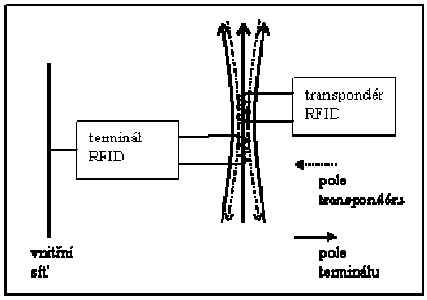
Frekvenční pásmo	Podtřída	Typický druh	Typické použití	Standardy	Operační vzdálenost (řádově)	Běžná bezpečnost
LF (100 až 150 kHz)	-	Paměťová karta	Přístupový systém, imobilizéry, implantáty, věrnostní karty.	proprietární	cm	L až M
HF (13.56 MHz)	Karta s vazbou na dálku	Paměťová karta	Přístupový systém, skipass, věrnostní karty.	ISO 15693	cm až m	L až M
	Karta s vazbou na blízko	Bezkontaktní smartkarta	Přístupový systém, platební karta, el. pas.	ISO 14443 ISO 7816	cm	M až H
UHF (800 MHz – 1 GHz)	-	Paměťová karta	Skladové hospodářství.	EPC [3] ISO 18000	cm až m	L až M

Tabulka 1: Klasifikace čipů pasivního RFID podle pracovní frekvence

3. Protokoly pásem LF a HF

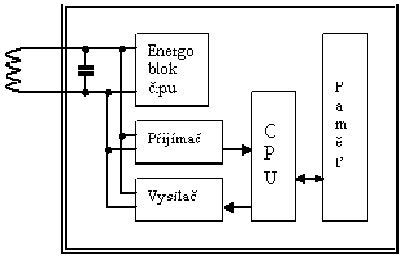
3.1 Fyzická vrstva

Podobně jako v případě mobilního telefonu, rozhlasu či klasické televize se k přenosu dat mezi terminálem a pasivním čipem RFID v pásmu LF a HF používá elektromagnetické pole. Vzhledem k právě uvedeným příkladům jsou zde ovšem dvě zásadní odlišnosti, které spolu úzce souvisí. První z nich je, že terminál do čipu nezasílá jen data, ale i energii pro jeho napájení. Právě proto a s ohledem na stále relativně nízké frekvence (nízké v tom smyslu, že pro popis elektronických obvodů ještě není nutné aplikovat složité modely s rozprostřenými parametry) je v centru pozornosti konstruktérů pouze magnetická složka pole indukovaná bezprostředně anténní cívkou terminálu či transpondéru. Na soustavu anténa terminálu – anténa čipu je zde nahlíženo zjednodušeně jako na vysokofrekvenční transformátor. To je druhá zásadní odlišnost oproti tomu, jak se s elektromagnetickým polem „zachází“ například v mobilním telefonu. Pro úplnost dodejme, že vlnový popis pole společně s obvody s rozprostřenými parametry přichází ke slovu u čipů v pásmu UHF.



Obrázek 1: Ilustrace indukční vazby terminálu a transpondéru RFID

V klidovém stavu, kdy neprobíhá komunikace ani jedním směrem, generuje terminál do své antény tzv. základní nosnou, což je harmonický signál o stabilní frekvenci. Pro pásmo LF to může být v zásadě libovolný kmitočet z rozsahu 100 až 150 kHz s tím, že často je volena hodnota 125 kHz. Pro pásmo HF je frekvence pevně stanovena příslušným standardem a činí 13.56 MHz. Díky indukční vazbě mezi oběma anténami budí základní nosná na anténě transpondéru určité napětí, ze kterého vnitřní energoblok odvozuje napájení celého čipu. Z teoretických analýz běžných typů antén vyplývá, že schopnost předávat dostatečné napájecí napětí klesá zhruba s faktorem $1/r^3$, kde r je vzdálenost obou antén [18]. Na první pohled nás může faktor $1/r^3$ zarazit, neboť útlum elektromagnetických vln se v radioelektronice běžně odhaduje faktorem $1/r^2$, který vychází z představy kulových vlnoploch. Hodnota $1/r^3$ je ovšem také v pořádku a je dána poněkud odlišným pohledem na využití pole k zajištění čistě indukční vazby obou antén. To se v důsledku podepisuje na operační vzdálenosti těchto čipů z nichž pouze karty s vazbou na dálku dokáží spolehlivě pracovat až na metrové vzdálenosti, čehož dosahují tím, že si musí dle standardu ISO 15693 povinně vystačit jen s desetinovou intenzitou pole terminálu oproti kartám s vazbou na blízko dle ISO 14443.



Obrázek 2: Vnitřní uspořádání transpondéru

Komunikace ve směru od terminálu k čipu probíhá tím způsobem, že terminál vhodně kódovaným datovým signálem moduluje základní nosnou, a to nejčastěji změnou její amplitudy (amplitudová modulace). Tím se mění i pole indukované anténou a tím i napětí sbírané na anténě transpondéru. Ten signál nejprve demoduluje vhodným detektorem a poté dekóduje. Dodejme, že blok přijímače čipu pracuje nezávisle na jeho energobloku, který musí být schopen i z modulované základní nosné sbírat dostupnou napájecí energii. Odpověď čipu je nejprve opět kódována a poté pomocí tzv. zátěžové modulace přenesena polem do terminálu. Při zátěžové modulaci čip určitým způsobem mění zatížení své antény (připojováním/odpojováním umělé zátěže), čímž generuje vlastní proměnné magnetické pole,

které se podle principu superpozice skládá s polem generovaným základní nosnou terminálu. Díky indukční vazbě mezi oběma anténami se toto celé projeví změnou napětí na anténní cívce terminálu. Do detailů zde zabíhat nebudeme, omezíme se na konstatování, že podle Lenzova pravidla [4] se zvýšení zátěže antény čipu projeví snížením napětí na anténě terminálu a naopak. Takový signál už potom terminál snadno demoduluje vhodným detektorem a dekóduje, čímž získá binární data zasílaná čipem. To vše opět probíhá nezávisle na energobloku transpondéru. Poznamenejme, že pokud zde byla řeč o kódování, tak tím bylo vždy míněno kódování za účelem kontroly chyb a efektivního využití přenosového pásma. Kryptografické ochrany do standardní výbavy na této úrovni bohužel nepatří.

3.2 Vyšší vrstvy pásma LF

U čipů pro pásmo LF většinou nějaké sofistikované protokoly vyšších vrstev nenajdeme. Vše je často navrženo velmi účelově tak, aby byly pokryty základní požadavky na předání několika jednoduchých příkazů a získání příslušných odpovědí. Transpondéry představující paměti s konstantním identifikačním řetězcem přenos dat od terminálu do čipu ani nepodporují, protože ho jednoduše nepotřebují. Jakmile se dostanou do správného pole, tak jen stále dokola pomocí zátěžové modulace vysílají svůj identifikátor. Kódování přenášených dat a případně i příkazové rámce (jsou-li podporovány) se čip od čipu liší, díky čemuž zde existují zhruba desítky schémat a protokolů. To pochopitelně komplikuje univerzálnost aplikací, kdy si konstruktér musí sám zavádět vhodné úrovně abstrakce a modulární architektury, aby se tak dopředu připravil na náhradu jednoho typu čipu za jiný.

3.3 Vyšší vrstvy pásma HF

Karty s vazbou na dálku

Elementární protokol a příkazy těchto karet popisuje standard ISO 15693. Základ pokrytý standardem se ovšem soustředí především na datové rozhraní a co se příkazů týká, tak pokrývá víceméně jen záběr odpovídající sofistikovanější paměťové kartě. Z bezpečnostních funkcí je zde zastoupena jediná, a to sice možnost uzamčení datových bloků do stavu pouze pro čtení. Vše ostatní je ponecháno na případném rozšíření o proprietární příkazy a služby čipů od jednotlivých výrobců. Na rozdíl od pásma LF však tato rozšíření alespoň staví na jasně daném základě, kterým je transportní protokol stanovený v uvedeném standardu. Na obrázku 3 vidíme typickou aplikaci těchto karet využívající jejich prodlouženou komunikační vzdálenost – skipass.

Obrázek 3: Skipass v pásmu HF podle ISO 15693

Karty s vazbou na blízko

Čipy z této kategorie jsou po protokolové a příkazové stránce bezpochyby nejvyspělejšími zástupci pasivních transpondérů RFID. Popisem rádiového rozhraní se zabývají standardy ISO 14443-1 až 3. Dodejme, že z jistých, zejména politicko-komerčních důvodů existují podstandardy označované jako 14443A a 14443B popisující čipy typu A, respektive typu B. Odlišnosti jsou přitom zejména v detailech rádiového kanálu. Na uvedené díly navazuje ISO 14443-4, který popisuje protokol pro předávání příkazů a odpovědí. Ten už je pro oba typy shodný, takže od této úrovně výše můžeme od jejich rozlišování abstrahovat. Samotný protokol zavedený ve čtvrtém dílu standardu je dále velmi podobný protokolu T=1 zavedenému dříve pro oblast kontaktních smartkaret standardem ISO 7816-3. Tato podobnost je nade vší pochybnost cílená, neboť je všeobecně předpokládáno, že díky ní dojde postupně k abstrakci komunikačního rozhraní moderních smartkaret. Transportní protokol zavedený v ISO 14443-4 je totiž v ISO 7816 zatím zpola neformálně zaveden pod označením T=CL (ContactLess) s tím, že od úrovně 7816-4 popisující konkrétní příkazy výše je už abstrahováno od toho, je-li komunikace s čipem kontaktní či bezkontaktní. V praxi už některé tzv. čtečky čipových karet nabízejí jak kontaktní tak i bezkontaktní rádiové rozhraní s tím, že bezkontaktní kanál vypadá z pohledu operačního systému jako další čtečka kontaktních karet schopná komunikovat protokolem ISO 7816-4 a vyšším. Dodejme, že hojně rozšířené rozhraní PC/SC [28] pro komunikaci s čipovými kartami takovou abstrakci samo aktivně podporuje. Aplikace napsané původně pro kontaktní smartkarty tak při přechodu ke stejně schopným kartám bezkontaktním ani nemusí zaznamenat nějakou změnu. Tento předpoklad byl v době psaní článku ověřen praktickým experimentem s aplikací napsanou původně pro kontaktní smartkarty s využitím rozhraní PC/SC v prostředí MS Windows XP. Jako čtečka byl použit model CardMan 5321 [26]. Vybraná aplikace nebyla nijak upravována, přesto byla schopna ihned bez problémů s bezkontaktní smartkartou přes kanál PC/SC a uvedený typ čtečky pracovat. Ilustraci uspořádání standardů ISO 14443 a ISO 7816 ukazuje tabulka 2. Čipy, které odpovídají celé této hierarchii, nazýváme bezkontaktní smartkarty. Podotkněme, že existují i účelové transpondéry, které propojení s ISO 7816 nepodporují. Příkladem může být platforma MIFARE zmíněná dále.

<table>
<tr><td>Aplikační vrstva</td><td colspan="3">ISO 7816-4 a vyšší</td></tr>
<tr><td>Transportní vrstva</td><td rowspan="3">ISO 7816-3</td><td colspan="2">ISO 14443-4</td></tr>
<tr><td>Linková vrstva</td><td>ISO 14443A-3</td><td>ISO 14443B-3</td></tr>
<tr><td>Fyzická vrstva</td><td>ISO 14443A-2</td><td>ISO 14443B-2</td></tr>
<tr><td>Elektromechanické vlastnosti</td><td>ISO 7816-1, 2</td><td colspan="2">ISO 14443-1</td></tr>
</table>

Tabulka 2: Uspořádání standardů pro kontaktní a bezkontaktní smartkarty

3.4 Pohled útočníka

Útočné vzdálenosti a techniky

Vyšetřujeme-li možné způsoby vedení útoku proti čipům RFID, otvírají se nám oproti například klasickým kontaktním smartkartám nové dimenze. Zatímco u kontaktních čipů začíná většina útoků předpokladem, že útočník nějakým způsobem dostal napadené zařízení

do rukou, v případě RFID tomu tak být rozhodně nemusí. V duchu celé technologie totiž i samotné útoky mohou být jaksi *bezkontaktní*. Z tohoto pohledu nás nyní jistě zajímá, na jakou vzdálenost lze o něčem takovém uvažovat. Pro začátek můžeme vyjít z operační vzdálenosti jednotlivých čipů uvedených v tabulce 1. Přiblíží-li se držitel čipu na méně než 10 cm k anténě útočníka, stává se potenciální obětí. Otázkou pochopitelně je, co mu může takový útočník provést. Pokud je čip a vůbec celá navazující aplikace řádně zabezpečená, tak nejspíš nic - snad jen zničit čip silným polem, jenže tak tomu nemusí být vždy. V příkladu uvedeném v části 4 například může útočník získat „unikátní" identifikátor, který si potom umístí do svého padělku (viz dále). Může se také pokusit o přepojovací útok, který ilustrujeme v části 6. Budeme-li se zabývat prakticky realizovatelnými scénáři útoků, zjistíme, že nabízející se možnosti jsou rozhodně zajímavé. Natolik zajímavé, že by stálo za úvahu snažit se běžnou operační vzdálenost prodloužit. Teoretické odhady a praktické experimenty [14] a [9] však ukazují, že zásadní vylepšení zde není reálné očekávat. Konkrétně v pásmu HF se u karet s vazbou na blízko dostaneme řádově na několik desítek cm.

Metoda	**Vzdálenost**
Aktivní komunikace s čipem	desítky cm
Pasivní odposlech – čip i terminál	jednotky m
Pasivní odposlech – jen terminál	desítky m
Aktivní komunikace s terminálem	desítky m

Tabulka 3: Útočné techniky a vzdálenosti pro pásmo HF dle ISO 14443

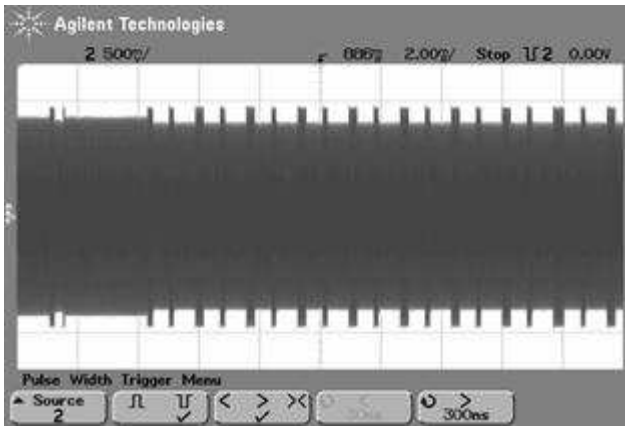
Až dosud jsme se ovšem zabývali případem, kdy útočník napadený čip musí i sám napájet. Říkejme tomu aktivní komunikace. Naproti tomu při pasivním odposlechu, kdy čip komunikuje s nějakým terminálem, který jej napájí, lze o významném prodloužení vzdálenosti uvažovat. Odhady pro pásmo HF uvádí tabulka 3, zde uvedené údaje vycházejí z veřejně dostupných studií [14], [5], [9]. Pro pásmo LF lze teoreticky očekávat podobné výsledky, neboť použité principy a metody lze do frekvenčního rozsahu LF přenést. Vidíme, že pokud nám bude stačit pouze slyšet stranu terminálu, může být útočník až několik desítek metrů daleko. Rapidní nárůst vzdálenosti je dán tím, že útočníkovo zařízení má vlastní zdroj energie, takže u signálu terminálu stačí jen úroveň nutná k jeho zachycení a demodulaci. Dodejme, že k některým, víceméně demonstračním útokům na elektronické pasy toto stačí [29], [16]. Za zmínku stojí i aktivní komunikace s terminálem, kdy útočníkovo zařízení předstírá, že je nějakým čipem RFID. Vidíme, že takové zařízení může být opět až několik desítek metrů daleko. To výrazně zjednodušuje praktickou realizaci padělků, jak si ukážeme v části 6.

Postranní kanály

Útoky založené na postranních kanálech [32] stále patří mezi nejúspěšnější nástroje moderní kryptoanalýzy. Uveďme, že postranním kanálem nazýváme každý nežádoucí způsob výměny informací mezi kryptografickým modulem (zde tedy čipem RFID) a jeho okolím. V případě hardwarových zařízení se jako nejúčinnější jeví kanály založené na sledování průběhu určitých fyzikálních veličin, jako je například spotřeba elektrické energie, elektromagnetické pole v okolí, atp. Možná protiopatření lze rozdělit do dvou základních skupin. Do první patří snahy o úpravy elektronické části tak, aby se výskyt daného kanálu eliminoval už v samotném počátku. Druhou skupinu tvoří techniky vycházející z úpravy samotných kryptoschémat, které

mají učinit unikající informaci neužitečnou. Teorie i praxe přitom jasně ukazují, že v chráněných zařízeních je vhodné oba přístupy kombinovat. V případě čipů RFID lze ovšem očekávat, že opatření prvního druhu budou velmi komplikovaná. Vzpomeneme-li si na způsob, jakým čip předává terminálu data pomocí zátěžové modulace, tak si snadno uvědomíme, že anténní obvod transponderu je a musí být cíleně konstruován tak, aby se změny v jeho zatížení zřetelně projevily v okolním elektromagnetickém poli. Jenže tou samou cestou se potom snadno odvysílají i změny způsobené závislostí spotřeby energie na probíhajícím kryptografickém výpočtu. Anténa totiž jen těžko „rozpozná“, co má propustit a co už ne. Taková filtrace je v principu možná v jednotce energobloku, avšak její praktická realizace bude nepochybně náročnější než u klasických kontaktních karet. Podotkněme, že ani u nich dosud tento problém není uspokojivě vyřešen.

Vidíme, že u čipů RFID bude v případě obrany proti postranním kanálům záležet zejména na protiopatřeních druhého druhu, tedy na vhodném uspořádání chráněného výpočtu. Pokud této otázce nebude věnována dostatečná pozornost, lze očekávat razantní útoky směřující k totálnímu prolomení, viz například [27], [10]. Pro ilustraci jisté nevyhnutelnosti existence postranních kanálů je na obrázku 4 uveden výsledek základního měření magnetického pole v okolí elektronického pasu ČR [19] v okamžiku výpočtu podepisovací transformace RSA [21]. Pro zvýraznění demonstrovaného postranního kanálu bylo nutné volit pomalejší časovou základnu, takže základní nosná o frekvenci 13.56 MHz se zde jeví jako souvislý horizontální pruh uprostřed obrazu. Vidíme, že amplituda nosné v jistých okamžicích zřetelně kolísá a vytváří na horním okraji jakési „cimbuří“. Totéž se pochopitelně projevuje symetricky na okraji spodním. Lze ukázat, že prohlubně zmíněného cimbuří s velkou pravděpodobností odpovídají okamžikům, kdy pracuje matematický koprocesor a hlavní procesor čeká na výsledek. Poměr délek navazujících prohlubní v určitých částech cimbuří přitom odpovídá poměru trvání dvou základních operací, které se při výpočtu transformace RSA vyskytují, a to sice mocnění (x^2 mod N) a násobení ($x*y$ mod N). Podle nich se celý postup výpočtu označuje termínem square-and-multiply. Pravděpodobně tedy umíme dost přesně určit, co v jaký okamžik matematický procesor dělá. Pokud by se jednalo o klasickou podobu uvedeného algoritmu, kterou najdeme v mnoha běžných implementacích RSA, byla by tím celá tzv. aktivní autentizace českých pasů prolomena! Naštěstí ale konstruktéři použili metodu zvanou square-and-multiply-always [10], která je proti přímočarým útokům postranními kanály odolná. Otázkou pochopitelně zůstává, jak přesně je tato metoda realizována a zda se přeci jen nenajde cesta k jejímu prolomení. První nájezd byl však úspěšně odražen.



Obrázek 4: Elektromagnetický postranní kanál elektronického pasu ČR [19]

4. Transpondéry unikátního ID

Dodnes se často setkáme s aplikacemi RFID, jejichž principem je transpondér nesoucí nějaký konstantní binární řetězec. Jakmile se takový čip dostane do pole terminálu, začne tuto hodnotu stále dokola vysílat, dokud pole terminálu zase nezmizí. Není zde přitom žádný protokol, který by nějak umožňoval kryptograficky ověřit, že předaná hodnota pochází z originálního čipu a ne z nějaké jeho kopie. Abychom si rizika takového přístupu ilustrovali na konkrétním příkladu, podíváme se zde blíže na transpondér typu EM4x02 [7], který můžeme často najít například coby klíč k domovním dveřím a garážím řady obytných domů v ČR.

Čip typu EM4x02 je určen pro pásmo LF. Z technického hlediska jde o sériovou paměť s kapacitou 64 bitů a konstantním obsahem. Obsah paměti a jeho interpretace zde není důležitá. Postačí uvést, že se jedná o binární posloupnost, kterou čip v poli čtečky cyklicky vysílá počínaje indexem 1 a konče pozicí 64. Pro rádiový přenos je obvykle použit kód Manchester a zátěžová modulace popsaná v části 3. Obecně pak přicházejí v úvahu ještě další dvě kódová schémata. Ani jedno z nich však nemá přenos kryptograficky chránit, takže se jimi zde zabývat nemusíme. Stačí konstatování, že jsou veřejně k dispozici [7]. O interpretaci vysílaných dat se čip nestará, to je věcí čtecího terminálu. Většina čteček se chová tak, že po kontrole integrity (nejedná se ovšem o kryptografickou integritu, účelem je jen detekce a případně i oprava chyb v přenosu) zpracuje přijatá data do podoby řetězce délky 5 B, který coby sériové číslo čipu předá na svůj výstup. Odtud si ho vyzvednou další komponenty systému (řízení zámků, atp.). Pro úplnost dodejme, že komunikační kanál ve směru čtečka → čip není implementován, neboť není zapotřebí.

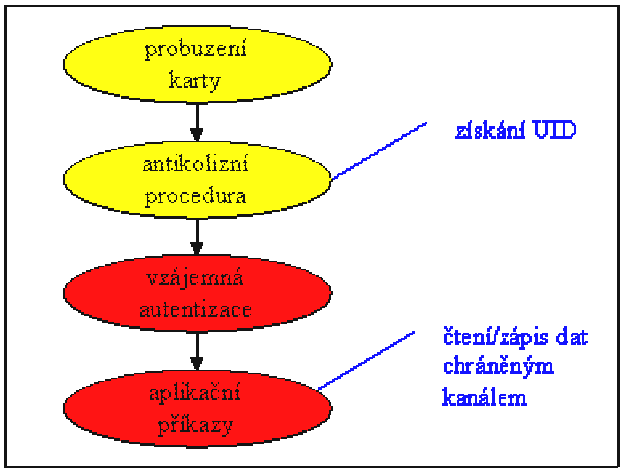
Obrázek 5: Transpondér v podobě klíčenky a zlodějka pro pásmo LF

Z uvedeného vyplývá, že pokud by například někdo chtěl vyrobit duplikát klíče od jisté garáže, postačí mu k tomu vědět, jaká binární posloupnost je v klíči uložena a mít k dispozici zařízení, které ji do příslušného terminálu přehraje. Ačkoliv se nás někdo může snažit přesvědčit, že konstrukce takového zařízení je vysoce komplikovaná a náročná záležitost, opak je pravdou. Na obrázku 5 vidíme zařízení sestavené volně dle stránek [33], které lze výstižně nazvat zlodějkou. Jak jinak také nazvat zařízení velikosti mobilního telefonu, které při přiblížení k výše popsanému typu čipu přečte a uloží obsah jeho paměti, aby ho potom kdykoliv později aktivně přehrálo do libovolné čtečky? Z elektronického hlediska přitom není překvapující, že něco takového lze sestrojit. Překvapující je, jak málo si toho jsou někteří aplikační konstruktéři vědomi. Pro úspěšné přečtení čísla transpondéru je nutné přiblížení na

zhruba centimetrovou vzdálenost. To je sice z pohledu útočníka jistá obstrukce, avšak v prostředcích MHD, ve frontě, atp. lze toto provést i bez vědomí jeho držitele. Vhodnou modifikaci zlodějky lze rovněž často snadno umístit a zamaskovat i do okolí řádného terminálu například u vjezdu do garáže. Při pasivním odposlechu lze uvažovat i o příslušném prodloužení vzdálenosti (experimentální výsledky bohužel chybí). Pak stačí jen přijít, vybrat si vhodnou zachycenou identitu a vstoupit...

5. Platforma MIFARE

Karty typu MIFARE [22] jsou dnes už spíš klasickým než přímo vzorovým příkladem HF čipů typu karta s vazbou na blízko (viz část 3). Jejich komunikační rozhraní je sice kompatibilní se standardem ISO 14443-A, avšak nad tímto rozhraním už nenajdeme pokračování dle ISO 7816, nýbrž příkazy proprietárního aplikačního protokolu MIFARE. Ten si společně s algoritmem Crypto1 firma Philips dobře hlídá. Objevily se ovšem už zprávy, že Crypto1 byl reverzním inženýrstvím odhalen a byla zahájena jeho intenzivní kryptoanalytická studie [24]. Kromě podnikových docházkových systémů se s technologií MIFARE setkáme například při placení za služby hromadné dopravy některých měst ČR.
V praxi se můžeme setkat s kartami MIFARE o kapacitě 1 KB nebo 4 KB. Dostupná paměť se dělí do bloků po 16 bajtech a tyto bloky se dále sdružují do takzvaných sektorů. Na rozdíl od čistě paměťových karet řídí MIFARE přístup k paměťovým blokům v rámci určitého sektoru na základě prokázání znalosti příslušného kryptografického klíče. Zde je použita varianta tříprůchodové autentizace dle ISO 9798-2, při níž se zároveň dohodnou klíče pro následné šifrování komunikace mezi kartou a terminálem. Dodejme však, že primární autentizační klíč má jen 48 bitů, takže zdrcujícímu útoku hrubou silou brání jen skutečnost, že kryptografický algoritmus s názvem Crypto1 je neveřejný. Jak jsme ale uvedli, tato ochranná bariera už možná padla [24]. Z toho vyplývá, že uvedený typ karet by už neměl být osazován do nových bezpečnostně kritických aplikací a ve starých by měl být postupně nahrazován vhodným typem bezkontaktní smartkarty.



Obrázek 6: Průběh sezení s kartou MIFARE

Například v přístupových systémech můžeme ovšem najít řadu aplikací MIFARE, které jsou a byly triviálně prolomitelné i bez objevů [24]. Jejich poznávacím znamením je, že z celé karty je zajímá jen a pouze její sériově číslo. Toto číslo je dostupné ihned během antikolizní procedury, takže z elektronického hlediska se čtecí terminál velmi zjednoduší, neboť není

nutné implementovat speciální obvody pro podporu transportního a aplikačního protokolu MIFARE (viz obrázek 6). Na druhou stranu se tím ovšem podstatně zjednoduší i případný útok, neboť tímto jsou schopnosti celé karty de facto degradovány na transpondér unikátního ID. Podobně jako pro pásmo LF lze i pro HF dle ISO 14443 sestavit vhodnou zlodějku a tu následně zneužít k vytváření duplikátů přístupových klíčů. Podíváme-li se podrobněji na to, jak vypadá antikolizní procedura dle ISO 14443, v rámci které se sériové číslo čipu předává, tak zjistíme, že rizika jsou dokonce ještě vyšší, než u zmíněných transpondérů v pásmu LF. Sériové číslo sice primárně vysílá čip, avšak na konci procedury jej terminál opakuje. Podíváme-li se na tabulku 3, je nárůst rizika zjevný – číslo opakované terminálem lze totiž zachytit až na desítky metrů daleko! Toto číslo je přitom kompletním klíčem, který po umístění do vhodné zlodějky odemkne příslušný zámek...
V souvislosti s MIFARE se také stává, že kdosi zapomene změnit tovární hodnoty přístupových klíčů. Výrobce s nimi sice dělá drahoty, ale uznejme, že implicitní řetězce A0A1A2A3A4A5 či B0B1B2B3B4B5, které lze volně najít na internetu, nejsou to pravé.

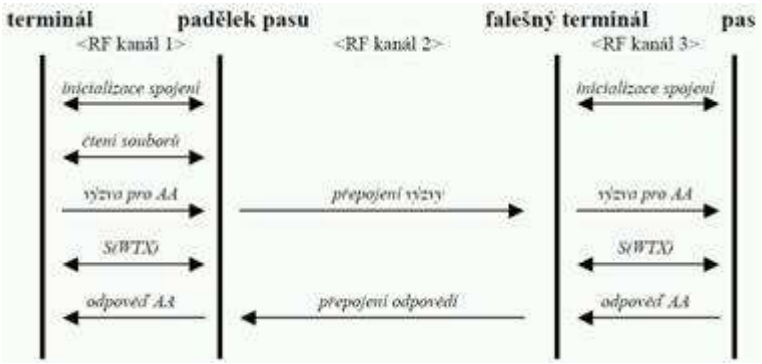
6. Elektronický cestovní pas

Elektronický cestovní pas ČR [31], [29], [30] vydávaný od polovinu roku 2006 obsahuje čip RFID pro pásmo HF. Konkrétně se jedná o typ karta s vazbou na blízko, který splňuje požadavky na bezkontaktní smartkartu (viz část 3). V porovnání s dříve zmíněnými platformami sériových pamětí pásma LF či MIFARE se nepochybně jedná o značně vyspělejší technologii. Mohlo by se zdát, že zde už žádné jednoduché útoky plynoucí z podstaty rádiového rozhraní nenajdeme, ale opak je pravdou. Jejich přehled je možné najít v publikacích [15], [16], [17], přičemž je vhodné dodat, že zatím nejde o zranitelnosti, které by elektronický pas z praktického hlediska významně oslabily. Lze na nich ale dobře ilustrovat, na co si musí návrhář aplikací pro bezkontaktní smartkarty dávat pozor. Zde se konkrétně podíváme na přepojovací útok na takzvanou aktivní autentizaci čipu elektronického pasu, která má za úkol bránit vytváření jejich neautorizovaných kopií.

Povinný digitální podpis dat uvedených v čipu elektronického pasu (tzv. pasivní autentizace) z principu nemůže zabránit vytvoření elektronicky identické kopie pasu. Stačí soubory podepsaných dat jednoduše překopírovat včetně souboru s jejich podpisem. Proto byl coby doplňkový mechanizmus navržen a v českých pasech (na rozdíl například od německých) i implementován jednoduchý protokol výzva-odpověď, který přímočarému kopírování brání. Německu se patrně nelíbila skutečnost, že záznam obsahující určitou specifickou výzvu (v otevřeném tvaru například „Bonn-A17 15.2.2008 17:20“) a správnou odpověď pasu může sloužit jako jistý druh důkazu přítomnosti pasu na nějakém místě. Za svou snahu chránit soukromí občanů však Německo zaplatilo silnou medializací „objevu“, že tamní pas lze okopírovat [25].

Podstatou diskutovaného protokolu je podpisové schéma RSA dle standardu ISO 9796-2 [21]. Veřejný klíč RSA je uložen v datovém souboru označovaném jako DG15, který si například terminál na letišti z předloženého pasu bez potíží přečte. Pro přehlednost výkladu zde vynecháváme problematiku autentizace terminálu pasu [29], [16], která z principu nemá na popsaný útok vliv. Soukromý klíč je uložen ve skrytém souboru, jehož externí čtení pas nedovoluje. Dále popsaný protokol umožňuje terminálu ověřit, že pas zná správný klíč do páru s klíčem z DG15, aniž by se přitom jeho hodnotu dozvěděl. Protokol zahajuje terminál 8B výzvou V. Pas vygeneruje náhodné číslo U v délce 106 bajtů a vypočte hašový kód w = SHA-1($U \parallel V$), který zformátuje do řetězce m = 6A $\parallel U \parallel w \parallel$ BC v délce 128 B (délka

modulu RSA u pasu v našich experimentech byla 1024 bitů). Na něj je aplikována podpisová transformace RSA a výsledek $s = m^d \bmod N$, kde N je modul a d soukromý exponent, je odeslán terminálu, který jej odpovídajícím způsobem ověří (detaily viz ISO 9796-2 a [21]).



Obrázek 7: Schéma přepojovacího útoku na elektronický pas

Pomineme-li postranní kanály, odolnost vůči nimž je zatím výrobním tajemstvím (viz 3.4.2), není v současné době znám matematický postup umožňující uvedený protokol prolomit. V případě RFID musíme ovšem vždy počítat s možností takzvaného přepojovacího útoku [8], [9], [13]. Analýza komunikačního protokolu navíc odhalila jistou technickou slabinu, která přepojovací útok ilustrovaný na obrázku 7 výrazně usnadňuje [12]. Jádrem slabiny je S-blok WTX transportního protokolu (viz ISO 14443-4), kterým pas žádá terminál o prodloužení doby čekání na odpověď. Pas v našich experimentech konkrétně požadoval prodloužení na 4949 ms, což byl 16násobek běžné doby. Pokud terminál tuto dobu nepotvrdil, pas spolupráci vždy ukončil. Předpokládáme proto, že i terminály například na hranicích musí S(WTX) korektně zpracovat a téměř pětisekundové čekání akceptovat. Náš pas byl přitom obvykle schopen odpovídat už za méně než 950 ms. Při zhruba hraniční intenzitě pole pak za méně než 1250 ms (zpomalením výpočtu si čip asi šetří energii). Vzniklou časovou rezervu může útočník využít k předání výzvy kanálem 2 do zařízení nazvaného falešný pas, které je v blízkosti originálu a požádá ho o správnou odpověď. Originál mu ji poskytne dostatečně rychle na to, aby byla kanálem 2 zpět přes padělek dopravena včas do terminálu. Ostatní požadavky padělek nepřepojuje, neboť je umí obsloužit lokálně (má kopie příslušných souborů). Hypoteticky by tak na řádný pas, který byl přes noc uložen v hotelové recepci, mohl kdosi překročit státní hranice, aniž by se originál pohnul z místa, kde si ho ráno nic netušící majitel vyzvedne.

Fyzika použitého rádiového rozhraní navíc nabízí zvládnutí i jinak nesnadného úkolu, že padělek musí mít stále vizuální podobu pasu a klasické ochranné prvky. Z tabulky 3 vidíme, že aktivní komunikace s terminálem je teoreticky možná až na desítky metrů daleko. Pas, který by útočník držel v ruce a který by podal hraniční kontrole, by pak vůbec nemusel nějaký čip obsahovat. Místo něho by totiž veškerou elektronickou komunikaci s terminálem na hranicích obstaralo z dostatečné vzdálenosti (například z auta či cestovního kufru) zcela jiné zařízení, které již podobu pasu mít rozhodně nemusí. To by pak obsloužilo i kanál 2 s falešným terminálem.

7. Trendy

Duální smartkarta

Již několik let jsou v oběhu karty, které disponují jak kontaktním rozhraním dle ISO 7816 tak i nějakým druhem rozhraní bezkontaktního. V drtivé většině případů jde ovšem o takzvané *hybridní* karty, ve kterých jsou fyzicky umístěny dva čipy, z nichž jeden představuje kontaktní smartkartu a druhý (skrytý v plastu) je transpondérem RFID. Tyto karty je nutné důsledně odlišovat od nastupující generace karet *duálních*, které jsou vybaveny pouze jedním čipem, ke kterému lze přistupovat přes dvě různé fyzické vrstvy – kontaktní i bezkontaktní (viz tabulka 2 v části 3.3.2). Výpočetní vybavenost takových čipů přitom plně odpovídá současným požadavkům na kvalitní smartkarty. Tyto karty lze s výhodou využít při postupné změně fyzické vrstvy karet používaných v nějaké podnikové infrastruktuře. Důvodem k přechodu k bezkontaktní komunikaci může být například nižší poruchovost, neboť rádiovému rozhraní nehrozí mechanické opotřebení kontaktů čteček. Lze proto očekávat, že význam kontaktního rozhraní bude postupně klesat ve prospěch rádiové komunikace.

Bezkontaktní platební karty

Masivnější rozvoj této technologie lze zatím pozorovat zejména v USA. Dlužno ovšem podotknout, že tam uvedená emise karet se po bezpečnostní stránce příliš nevydařila [11]. Ačkoliv jsou americké karty založeny na standardu ISO 14443, nezdá se, že by propojení se standardem ISO 7816 směrem k bezkontaktní smartkartě nějak užitečně využívaly ke zvýšení bezpečnosti. Podle studie [11] karty disponují proprietárními příkazy, které mají za cíl hlavně učinit jejich zpracování v terminálech obchodníků co nejvíce podobným klasickým kartám s magnetickým proužkem dle standardu ISO 7813. Autoři ukazují na celou řadu prakticky schůdných útoků včetně přepojení relace či duplikace karty bez vědomí držitele. Nezbývá než doufat, že novější typy karet, které se již karetní asociace snaží podchytit podobným standardem jako klasické (kontaktní) čipové karty, se již vydaří lépe [2]. Platební karty jsou jistě rovněž zajímavou příležitostí pro duální karty zmíněné výše.

NFC

Obecným cílem rozhraní NFC (Near Field Communication) podle standardu ISO 18092 je zpřístupnit rádiové pásmo HF čipů RFID pro komunikaci několika zařízení vybavených vlastními zdroji elektrické energie. Mezi zmíněná zařízení by typicky měly patřit například mobilní telefony. Na první pohled by se mohlo zdát, že zrovna u mobilního telefonu je toto rozhraní poněkud zbytečným konkurentem k existující a zaběhnuté technologii Bluetooth [1]. Jenže NFC nabízí jednu zásadní novou vlastnost a tou je možnost emulovat v podstatě jakýkoliv transpondér dle standardu ISO 14443. Z příslušně vybaveného mobilního telefonu se tak může v případě potřeby stát bezkontaktní platební karta, vstupenka na koncert či věrnostní karta, atp. Emulujícím zařízením přitom nemusí být nutně jen mobilní telefon, lze uvažovat i o hodinkách, palmtopech atp. Praktické implementace NFC však zatím za jeho teoretickými možnostmi poněkud pokulhávají. Lze se domnívat, že trh čeká na rozšíření vhodné bezkontaktní aplikace, jejíž nosiče poté budou zařízení vybavená rozhraním NFC ke spokojenosti svých majitelů emulovat. Jednou z nich by mohl například často medializovaný projekt [20].

Je patrné, že zařízení vybavená rozhraním NFC, která jsou dostatečně flexibilně programovatelná, by se mohla zároveň stát vítanými nástroji útočníků. Dost možná se

v souvislosti s tím objeví snahy flexibilitu implementací NFC jaksi násilně omezovat, ačkoliv je pochopitelné, že tudy cesta do bezpečí určitě nevede.

8. Závěr

Představili jsme si platformu pasivního RFID pracujícího na frekvencích v pásmech LF a HF. Na konkrétních příkladech jsme si ukázali vlastnosti nejčastěji používaných čipů a upozornili přitom na možné hrozby jejich nesprávného aplikování. Určitě nelze tvrdit, že RFID s sebou musí vždy přinést oslabení systému, do kterého je nasazeno. To, že v praxi taková oslabení často pozorujeme, je zejména důsledkem planých očekávání a chybných předpokladů. Nároky, které správná aplikace těchto čipů na konstruktéra klade, totiž nejsou jen technického rázu, kdy je nutné do detailu pochopit, jak vše pracuje. Neméně důležitá je změna zažitého paradigma, kdy je nutné důsledně revidovat vzorce nacvičené používáním klasických kontaktních čipů. Za všechny zdůrazněme ten nejzásadnější – to, že terminál právě komunikuje s bezkontaktním čipem vůbec neznamená, že ho jeho řádný držitel vůbec vyndal z kapsy a že o tom celém ví. Tyto, řekněme, elementární zranitelnosti jsou dány specifickými vlastnosti použitého rádiového rozhraní. Zajímáme-li se o analýzu rizik spojených s RFID, určitě se nám proto vyplatí věnovat potřebný čas důkladnému pochopení jeho fyzické komunikační vrstvy.

Ačkoliv architektura MIFARE či snad dokonce i transpondéry unikátního ID budou asi ještě nějakou dobu dožívat ve stávajících aplikacích, trend směřuje k vyspělým čipům z kategorie karet s vazbou na blízko v pásmu HF. Konkrétně se jedná o duální čipové karty, které nabízejí k jednomu a témuž čipu jak kontaktní tak i bezkontaktní přístup. Lze si proto docela dobře představit dobu, kdy se budeme ke dveřím pracovny hlásit stejně robustními metodami jako k pracovní stanici podnikové sítě. Masovější penetraci této technologie může také výrazně napomoci rozhraní NFC umožňující v dobrém slova smyslu napodobit například bezkontaktní platební kartu mobilním telefonem či hodinkami. Upozorněme však že kromě zjevných výhod přináší NFC i nové specifické hrozby, které si vynucují další změnu zažitých návrhových předpokladů.

Podotkněme, že v tomto pojednání bylo záměrně ponecháno stranou téma ochrany soukromí versus RFID, které bývá v posledních letech zmiňováno tak často, že by snadno mohl vzniknout dojem, že o jiné problémy v souvislosti s RFID nejde. To by ovšem byl zásadní omyl, neboť ochrana soukromí je jen jedním z mnoha aspektů, které lze sledovat v souvislosti s nasazením této technologie. K tomu je ovšem nejprve nutné ji dobře poznat po technické stránce a uvědomit si její přirozené zranitelnosti, na základě kterých je pak následně možné analyzovat rizika toho kterého způsobu použití. Zde jsme se věnovali zejména první části s jistým akcentem pro oblast přístupových a identifikačních systémů, na které jako by se při „honech na čarodějnice“ v souvislosti se soukromím občanů úplně zapomnělo.

Reference

[0] Hlaváč, M., Rosa, T.: RFID: Co to vlastně máme v kapse? , *Information Security Summit, 9th International Conference*, Praha , pp. 29-40, 2008.

[1] Bluetooth, *Wikipedia*, http://en.wikipedia.org/wiki/Bluetooth .

[2] EMV Contactless Specifications for Payment Systems, http://www.emvco.com/specifications.asp?show=58.

[3] EPC Class-1 Generation-2 UHF RFID Conformance Requirements Specification v. 1.0.2.

[4] Feynman, R.-P., Leighton, R.-B., and Sands, M.: Feynmanovy přednášky z fyziky s řešenými příklady, California Institute of Technology, USA, 1964, Fragment, Havlíčkův Brod, 2001.

[5] Finke, T., Kelter, H.: Abhörmöglichkeiten der Kommunikation zwischen Lesegerät und Transponder am Beispiel eines ISO14443-Systems, *BSI - German Federal Office for Information Security*, 2005.

[6] Finkenzeller, K.: RFID Handbook – Fundamentals and Applications in Contactless Smart Cards and Identification, John Willey and Sons Ltd., 2003.

[7] H4102 – Read Only Contactless Identification Device, EM Microelectronic-Marin SA, SWATCH Group, 2000.

[8] Hancke, G.-P.: A Practical Relay Attack on ISO 14443 Proximity Cards, http://www.cl.cam.ac.uk/~gh275/relay.pdf , 2005 .

[9] Hancke, G.-P.: Practical Attacks on Proximity Identification Systems (Short Paper), in *Proc. of the 2006 IEEE Symposium on Security and Privacy (S&P'06)*, Vol. 00, pp. 328-333, 2006.

[10] Handschuh, H.: Contactless Technology Security Issues, *Information Security Bulletin*, Vol. 9, pp. 95 - 100, April 2004.

[11] Heydt-Benjamin, T.-S., Bailey, D.-V., Fu, K., Juels, A., and O'Hare, T.: Vulnerabilities in First-Generation RFID-Enabled Credit Cards, in Proc. of *Eleventh International Conference on Financial Cryptography and Data Security*, 2007.

[12] Hlaváč, M., and Rosa, T.: A Note on the Relay Attacks on e-passports: The Case of Czech e-passports, *IACR ePrint Report 2007/244*, 2007.

[13] Kfir, Z., Wool, A.: Picking Virtual Pockets using Relay Attacks on Contactless Smartcard Systems, *IACR ePrint Report 2005/052*, 2005.

[14] Kirschenbaum, I., Wool, A.: How to Build a Low-Cost, Extended-Range RFID Skimmer, in *Proc. of the 15th conference on USENIX Security Symposium*, pages 4 – 4, 2006.

[15] Klíma, V., a Rosa, T.: Elektronický cestovní pas – Komunikační rozhraní, *Sdělovací technika*, 2, 2007.

[16] Klíma, V., a Rosa, T.: Elektronický cestovní pas – Základní řízení přístupu, *Sdělovací technika*, 3, 2007.

[17] Klíma, V., a Rosa, T.: Elektronický cestovní pas – Autentizace, *Sdělovací technika*, 4, 2007.

[18] Lee, Y.: Antenna Circuit Design for RFID Applications, *Application Note 710*, Microchip Tech. Inc., 2003.

[19] Lórencz, R., Buček, J., a Zahradnický, T.: *osobní komunikace*, 2007.

[20] Mastercard PayPass - Documentation,

http://www.paypass.com/documentation.html#Technical%20Specifications .

[21] Menezes, A.-J., van Oorschot, P.-C., and Vanstone, S.-A.: Handbook of Applied Cryptography, CRC Press, 1996.

[22] MIFARE MF1 IC S50, Philips Semiconductors, Rev. 5.1, May 2005.

[23] New RFID chips to support Mastercard's PayPass cards,

http://rfdesign.com/news/RFID-Mastercard-Paypass/ , 2005.

[24] Nohl, K, and Plötz, H.: MIFARE – Little Security, Despite Obscurity, *24th Chaos Communication Congress*, 2007.

[25] O'Connor, M.-C.: Industry Group Says E-Passport Clone Poses Little Risk, *RFID Journal News*, Aug. 9, 2006, http://www.rfidjournal.com/article/view/2559/ .

[26] Omnikey CardMan 5312:

http://omnikey.aaitg.com/?id=products&tx_okprod_pi1%5Bproduct%5D=41 .

[27] Oren, Y., and Shamir, A.: Power Analysis of RFID Tags,

http://www.wisdom.weizmann.ac.il/~yossio/rfid/ .

[28] PC/SC Workgroup Specifications Overview,

http://www.pcscworkgroup.com/specifications/overview.php .

[29] Říha, Z.: Bezpečnost elektronických pasů - část I, *Data Security Management*, Praha : Tate International, Vol. XI, 1, pp. 26-31, 2007.

[30] Říha, Z., Švenda, P., a Matyáš, V.: Bezpečnost elektronických pasů - část II, *Data Security Management*, Praha : Tate International, Vol. XI, 2, pp. 46-49, 2007.

[31] Říha, Z., a Vakalis, I.: Elektronické pasy, *Data Security Management*, Praha: Tate International, Vol. X, 3, pp. 28-34, 2006.

[32] Side channel attack, *Wikipedia*, http://en.wikipedia.org/wiki/Side_channel_attack .

[33] Westhues, J.: RFID devices at http://cq.cx/vchdiy.pl .

B. Bezpečnost PHP aplikací

Jakub Vrána, jakub@vrana.cz, http://www.vrana.cz/

Na zabezpečení webových aplikací je potřeba dbát ještě více než na zabezpečení aplikací desktopových. Důvodem je to, že s webovým serverem se komunikuje pomocí protokolů, ve kterých se dají snadno měnit data. Druhým důvodem je, že ve výsledné stránce se kombinuje textový výstup s logikou stránky zapsanou pomocí značek.

Článek představuje několik základních útoků, které jsou proti webovým aplikacím vedeny, se zaměřením na stránky vytvořené v PHP.

1. Cross Site Scripting

Útok spočívá v tom, že se útočníkovi podaří na naši stránku umístit kromě textu také značky, které ovlivňují její chování. Nejčastěji se jedná o značku <script>, pomocí které může útočník v kontextu naší aplikace spustit vlastní klientský skript. Tento skript se dá použít např. ke zcizení tzv. session identifikátoru, který se používá pro jednoznačnou identifikaci uživatele naší aplikace, takže se za něj útočník může následně začít vydávat. Použití značky <script> ale není jediným rizikem tohoto útoku, stejně nebezpečné může být vložení falešného přihlašovacího formuláře, který pošle vyplněná data na server útočníka, nebo i prostá změna vzhledu stránky (např. umístění posměšného obrázku).

Útokem se musíme zabývat všude tam, kde do stránky vkládáme nedůvěryhodná data získaná obvykle jako vstup od uživatele. Typickým místem náchylným k útoku XSS jsou diskusní fóra, ale např. také vyhledávání nebo registrační formuláře zobrazující vyplněné hodnoty.
Obrana proti tomuto útoku je poměrně jednoduchá – stačí veškerá nedůvěryhodná data vkládaná do stránky ošetřit funkcí htmlspecialchars, která převede znaky se speciálním významem v HTML na odpovídající entity. Naopak nevhodná je funkce strip_tags, která odstraní všechny značky, protože svévolně modifikuje data zadaná uživatelem (někdo skutečně může chtít napsat kód značky) a navíc nijak neošetřuje znak ampersand, který má v HTML také speciální význam. Pokud chceme uživateli dát možnost nějak formátovat vstup, je nejlepší k tomu využít specializovaného nástroje, např. Texy! (http://texy.info/).

Funkci htmlspecialchars je vhodné aplikovat až při vypisování textu do stránky a nikoliv při ukládání dat do databáze. Díky tomu budou data v databázi uložena v čisté podobě, takže je budou moci použít i jiné aplikace, a na druhou stranu máme jistotu, že ať už bude v databázi

cokoliv, na výstupu dostaneme vždy bezpečný kód. V neposlední řadě jsou ošetřená data delší, takže by se do databázových políček nemusela vejít.

2. SQL Injection

Útok dovoluje útočníkovi pozměnit existující nebo vykonat vlastní SQL příkaz nad naší databází. Realizuje se tak, že útočník někde vloží místo čísla řetězec nebo že v řetězci použije speciální znaky jako např. apostrof. Kupříkladu za identifikátor zobrazeného článku přenášený v URL vloží kód UNION SELECT login, heslo FROM uzivatele, čímž získá přihlašovací údaje všech uživatelů na serveru. Nebo na místo hesla vloží kód ' OR 1=1 #, čímž ověření hesla zcela obejde a tím pádem se přihlásí jako libovolný uživatel.

K útoku SQL injection může dojít všude tam, kde se přímo do SQL dotazu vkládají nedůvěryhodná data. Zcela zabránit tomuto typu útoku lze využitím konceptu vázání proměnných, kdy se databázovému serveru posílá odděleně kód SQL dotazu a data, která tento dotaz používá. Např. v rozšíření PDO se k tomu používají metody PDO::prepare a PDOStatement::execute:

```
<?php
$result = $pdo->prepare("SELECT * FROM uzivatele WHERE login LIKE ?");
$result->execute(array('franta'));
?>
```

I bez vázání proměnných lze útoku poměrně snadno zabránit. Stačí veškerá nedůvěryhodná data v SQL příkazu uzavřít do apostrofů a ošetřit je odpovídající funkcí. Např. při použití rozšíření MySQL jde o funkci mysql_real_escape_string, v jiných rozšířeních existuje obdobná funkce. V MySQL lze při použití běžných kódování (jednobajtová nebo UTF-8) použít i obecnou funkci addslashes.

To je také funkce, která se automaticky aplikuje na všechna data od uživatele, pokud je zapnutá konfigurační direktiva magic_quotes_gpc. V tom případě lze do apostrofů v SQL dotazu beztrestně vkládat proměnné $_GET, $_POST a $_COOKIE, naopak je ale potřeba tato data odošetřit funkcí stripslashes při jiném použití (např. při přímém vypsání do stránky). Čísla, která v SQL nelze uzavřít do apostrofů (např. v klauzuli LIMIT), je možné ošetřit funkcí intval.

3. Podstrčení proměnných

PHP dovoluje pracovat s neinicializovanými proměnnými. Vygeneruje se při tom chyba úrovně E_NOTICE, ta je ale často ignorovaná nebo k ní dojde až v momentě zdárného útoku, kdy je na její ošetření už pozdě.

Útok podstrčení proměnných spočívá v tom, že se útočníkovi podaří nastavit hodnotu neinicializované proměnné. Nejčastěji k tomu může dojít při zapnuté konfigurační direktivě register_globals, která způsobí, že všechna data od uživatele budou k dispozici také v globálních proměnných. Druhá možnost podstrčení proměnné je z knihovny vkládané skriptem, k té ale útočník obvykle nemá přístup.

Skript náchylný k podstrčení proměnných vypadá typicky takto:

```
<?php
if ($_POST["heslo"] == "tajne_heslo") {
  $overeno = true;
}
if ($overeno) {
  // ...
}
?>
```

Proměnná $overeno není správně inicializovaná, takže pokud je zapnutá direktiva register_globals a útočník do adresního řádku přidá ?overeno=1, dostane se do zabezpečené části skriptu i bez znalosti hesla.

Spolehlivou obranou proti podstrčení proměnných je důsledná inicializace všech proměnných. Vhodným doplňkem je vypnutí konfigurační direktivy register_globals a prověřování chyb úrovně E_NOTICE, samo o sobě to ale bezpečnost nezaručí.

K útoku jsou nejvíce náchylné aplikace s otevřeným zdrojovým kódem, které mohou běžet na různých konfiguracích. U těch může útočník snadno zjistit názvy neinicializovaných proměnných a útok provést na konfiguraci, která ho umožňuje provést vzdáleně.

4. Vzdálené spuštění

Při vzdáleném spuštění se útočníkovi podaří spustit vlastní PHP skript na našem webovém serveru v kontextu naší aplikace, takže jde asi o nejnebezpečnější útok, protože útočník může např. získat a následně smazat všechna data v databázi a zdrojové kódy aplikace.

Typickou ukázkou náchylnou ke vzdálenému spuštění je tento kód:

```
<?php
include "$_GET[page].php";
?>
```

Díky direktivě allow_url_fopen, která je ve výchozím nastavení zapnutá, dovoluje PHP pracovat i se vzdálenými soubory. V PHP 5.2.0 byla zavedena nová direktiva allow_url_include, která ovlivňuje práci s vkládanými soubory a která je ve výchozím nastavení vypnutá a obvykle není důvod ji zapínat. Vzdálené spuštění je ale možné provést tak, že útočník na server nejprve nahraje skript a tento skript následně spustí lokálně, takže vypnutí těchto direktiv nás od útoku spolehlivě neuchrání. Pokud naše aplikace dovoluje uživatelům např. nahrávat fotky, může útočník skript schovat do nahraného souboru. Pokud aplikace běží na sdíleném webhostingu, může útočník nahrát skript do vlastního prostoru na serveru (následnému vložení takového skriptu se nicméně dá zamezit vhodnou konfigurací serveru). Spolehlivou obranou proti této variantě útoku je vyhnout se dynamickému vkládání skriptů na základě dat předaných od uživatele.

Uvedená ukázka není jediným zástupcem aplikací náchylných ke vzdálenému spuštění. Pokud dovoluje naše aplikace nahrávat uživatelům soubory ukládané do souboru na disku a následně přímo přístupné z prohlížeče, může útočník umístit svůj skript do tohoto souboru. Řešením

tohoto rizika je vypnutí direktivy engine, která dovoluje zpracování PHP skriptů, v adresáři s datovými soubory. Naopak nevhodnou obranou by bylo např. hledání posloupnosti <? v nahraných souborech (ta se může vyskytovat i jinde než v PHP skriptech) nebo kontrola koncovky nahraného souboru (protože webový server může PHP skripty zpracovávat i v souborech s netypickými koncovkami).

5. Cross-Site Request Forgery

Útok CSRF je netypický tím, že útočník vlastně neútočí na náš server, ale místo toho nechává oprávněného uživatele provádět operace podle svého uvážení. Z pohledu naší aplikace se jedná o legitimní pokyny oprávněného uživatele, ten je ale provádí nedobrovolně.

Představme si, že smazání záznamu v naší aplikaci realizuje skript zaznam.php?smazat=ID. Tento skript nejprve ověří, zda je uživatel přihlášen a zda má k mazanému záznamu příslušné oprávnění. Teprve v tom případě záznam smaže. Útočník ale udělá to, že tento odkaz umístí na vlastní nebo nějaké veřejné stránky (např. do atributu <img src="">, takže se při návštěvě stránky skript automaticky spustí) a počká, až na ně oprávněný uživatel přijde (nebo ho může k návštěvě nalákat).

Útok je založen na tom, že uživatel je ve webových aplikacích nejčastěji identifikován pomocí cookie, která se posílá transparentně bez vědomí uživatele. Útok by nebylo možno realizovat v situaci, kdy by se session identifikátor předával v URL, v tom případě by ale bylo potřeba toto URL důkladně chránit, aby se např. pomocí hlavičky Referer nedostalo na cizí servery.

Útok je možné provést i v případě, kdy se operace neprovádí metodou GET, ale správně metodou POST. Útočníkovi v tom případě stačí vytvořit na svých stránkách formulář s cílem na našem serveru a tento formulář automaticky odeslat klientským skriptem.

Spolehlivá obrana proti tomuto útoku spočívá ve vygenerování náhodného řetězce, tzv. tokenu na stránce, která předchází provedení operace (např. smazání záznamu předchází jeho zobrazení). Tento token se následně uloží do session proměnné a zároveň se pošle ve skrytém formulářovém poli. Před provedením operace se porovná hodnota tokenu v session proměnné a ve skrytém formulářovém poli a operace se provede jen tehdy, když jsou tyto hodnoty totožné. Útočník nemá hodnotu tokenu jak zjistit.

Útok se dobře provádí u veřejných aplikací, do kterých se může útočník sám přihlásit a prozkoumat jejich strukturu nebo které jsou někde popsané. Jinak musí adresy a názvy parametrů provádějících požadované operace uhodnout.

6. Session Hijacking

Při útoku Session Hijacking se podaří útočníkovi získat session identifikátor uživatele a následně se za něj začít vydávat. Tento identifikátor může útočník získat např. již představeným útokem XSS nebo ho může získat přímo na serveru – např. pokud se session data na sdíleném webhostingu ukládají do společného adresáře /tmp. Session identifikátor by útočník mohl získat i po cestě mezi serverem a klientem (nejčastěji v lokální síti klienta),

tomu lze zabránit použitím protokolu HTTPS, nebo ve špatně zabezpečeném prohlížeči uživatele.

Útoku Session Hijacking se brání poměrně těžko, protože uživatele nemáme kromě session identifikátoru téměř podle čeho poznat. Jeho IP adresa se může v čase měnit, naopak stejnou adresu může mít více uživatelů. Spolehnout se můžeme na některé hlavičky posílané klientem, např. User-Agent, to útočníkovi ale moc práce nepřidělá. Obranou proti tomuto útoku je tedy ošetřit všechna místa, kde by k úniku session identifikátoru mohlo dojít: XSS, server, HTTPS, klient a URL, pokud se session identifikátor přenáší v něm (čemuž je lepší se vyhnout).

7. Session Fixation

Útok Session Fixation je podobný jako Session Hijacking, ale útočník se nesnaží získat session identifikátor uživatele, ale naopak mu vnutí svůj. Následně počká, až se uživatel přihlásí, a začne se za něj vydávat. Obrana proti tomuto útoku je jednoduchá, stačí v momentě přihlášení změnit session identifikátor funkcí session_regenerate_id.

8. Response Splitting

Útok spočívá v modifikaci hlaviček přenosových protokolů, ať už HTTP u webových stránek nebo SMTP u e-mailů.

U webových stránek se tímto útokem dá způsobit Session Fixation nebo kompletně změnit obsah stránky a umístit na ni třeba falešný přihlašovací formulář. V aktuálních verzích PHP je tento útok nicméně těžko proveditelný, protože PHP od verze 4.4.2 a 5.1.2 nedovoluje ve funkci header poslat více než jednu hlavičku. I tak je ale vhodné se vyhnout předávání dat získaných od uživatele této funkci.

U e-mailů se Response Splitting používá především k rozesílání spamu. Pokud naše aplikace obsahuje formulář, který odesílá e-maily a dovoluje vyplnit adresu odesílatele nebo příjemce, předmět nebo další hlavičky (byť by se předávaly jen ve skrytém formulářovém poli), je potřeba se tímto útokem zabývat. Pokud totiž útočník např. do pole odesílatele vyplní "spam@example.com\nBcc: seznam-adres\n\nZde je spam.", vygeneruje se následující zpráva:

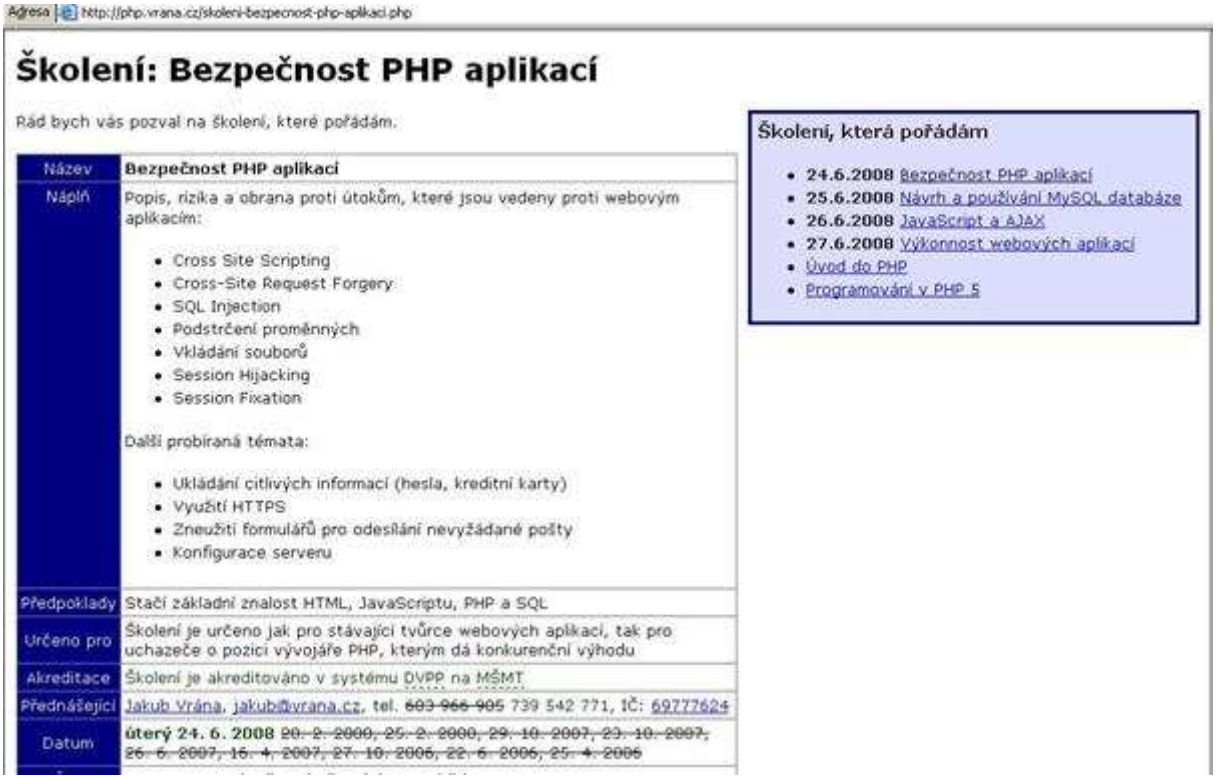
```
From: spam@example.com
Bcc: seznam-adres

Zde je spam.
```

Prázdný řádek se v protokolu SMTP používá pro oddělení hlaviček od těla zprávy, takže vše, co mu následuje, je chápáno jako zpráva. Tím může útočník pomocí podvržení hlavičky sestavit vlastní zprávu a rozeslat ji na vlastní seznam adres.
Obrana proti tomuto útoku spočívá v kontrole hodnot předávaných formulářem, zda neobsahují znak konce řádku.

9. Závěr

Obrana proti popsaným útokům je jen jednou částí komplexní úlohy zabezpečení webové aplikace. Další důležité oblasti zahrnují správné zpracování dat, ukládání citlivých informací, zabezpečení komunikace nebo nastavení serveru. Je potřeba se zabývat i různými způsoby kontroly bezpečnosti aplikace.



Přijďte si o tom všem včetně detailního rozboru a ukázky zmíněných útoků popovídat v úterý 24. 6. 2008 do Brna (nebo v některém dalším termínu) na jednodenní školení o bezpečnosti PHP aplikací (http://php.vrana.cz/skoleni-bezpecnost-php-aplikaci.php) , které zde pořádá autor tohoto článku, dlouhodobý spolupracovník e-zinu Crypto-World.

C. Popis šifrovacího algoritmu Serpent

Jan Jeřábek, Ústav telekomunikací, FEKT VUT v Brně, (jerabekj@feec.vutbr.cz)

Abstrakt. Tento článek připomíná dnes méně známý a přesto velmi kvalitní šifrovací algoritmus Serpent. Jedná se o symetrickou blokovou šifru, která má mnohé společné se standardem AES (algoritmus Rijndael), jelikož Serpent taktéž usiloval o vítězství v soutěži o novou normu na poli šifrování. Srovnání těchto dvou mechanizmů z obecného a výkonnostního hlediska a z pohledu hardwarové implementace je součástí textu, nejvýznamnějším rozdílem mezi nimi je počet hlavních rund při šifrování/dešifrování, který má podle předpokladů vliv na rychlost vykonávání těchto činností.

1. Úvod

V roce 1997 vyhlásila americká vládní organizace NIST (*National Institut of Standards and Technology*) [1] veřejnou soutěž spočívající ve vývoji a následném výběru nového standardu pro šifrování. Tento standard pro symetrické šifrování měl být pojmenován AES (*Advanced Encryption Standard*) a jeho úkolem bylo nahradit stávající a již nedostačující mechanizmus – DES (*Data Encryption Standard*). DES [2] pokulhával především poměrně malou délkou klíče (56 bitů) a částečně také tím, že byl navržen především pro hardwarovou implementaci a jeho softwarová provedení byla dosti neefektivní. Do soutěže se přihlásilo mnoho různých týmů, do finále postoupilo pět pravděpodobně nejlepších algoritmů – MARS [3], RC6 [4], Rijndael [5], Serpent [6] a Twofish [7]. Je poměrně dobře známo, že soutěž nakonec vyhrál algoritmus Rijndael a po dohromady skoro pětiletém procesu se stal v roce 2001 standardem AES. Tento algoritmus, který vyvinuli dva Belgičané – Joan Daemon a Vincent Rijmen, se dnes běžně využívá v mnoha aplikacích a je i po více než sedmi letech provozu stále považován za bezpečný, přestože jsou již známy některé méně závažné útoky [8]. Algoritmus Rijndael však soutěž o AES nevyhrál se zcela zásadní převahou, výsledky rozhodujícího hlasování byly (po započtení záporných hlasů) [9] následující: Rijndael 76, Serpent 52, Twofish 10, RC6 -14 a MARS dokonce -70. Výsledek vedl také k tomu, že zejména druhý a třetí algoritmus v pořadí jsou dnes také poměrně dost rozšířené. Tento článek je věnován druhému z výsledného pořadí soutěže – šifře Serpent. Např. podle testu na bezpečnost, který provedli autoři algoritmu Twofish [10], je totiž Serpent výrazně bezpečnější než Rijndael.

2. Popis algoritmu

2.1 Obecný popis šifrovacího procesu

Popis algoritmu je možné nalézt především v [6]. Z textu je patrné, že autoři k návrhu šifry přistoupili velmi konzervativně, snažili se používat pouze dobře prověřené komponenty a vyhýbat se nedostatečně známým řešením, což by mělo zaručit, že šifra nebudou obsahovat žádná „zadní vrátka“. Při návrhu vyšli z algoritmu DES, ale bylo samozřejmě nutné provést mnoho změn, aby Serpent vyhovoval požadavkům na AES.

Serpent je složen z celkem 32 rund substitučně-permutační sítě pracujících se čtyřmi 32-bitovými slovy, výstupem je tedy blok o délce 128 bitů. Každým 128 bitům otevřeného textu na vstupu algoritmu je tedy na výstupu přiřazen 128 bitů dlouhý kryptogram. Šifrovací klíč uživatele šifry může být libovolné délky až do maximální délky, 256 bitů. Při šifrování

(resp. pro tvorbu rundovních klíčů) se však vždy používá 256 bitový klíč, nehledě na to, jak dlouhý klíč uživatel zadal. Toho je docíleno primitivní operací – doplnění uživatelského klíče hodnotou „1“ na pozici MSB a následně odpovídajícím počtem bitů s hodnotou „0“, tak aby celý klíč dosáhl požadovaných 256 bitů.

Počet rund (32) je zvolen poměrně značně vysoký. Autoři algoritmu tvrdí, že kdyby se počet rund snížil na polovinu (16), Serpent by poskytoval stejnou úroveň kryptografického zabezpečení jako triple-DES. Použitím dvojnásobného počtu rund je podle autorů možné dosáhnout vynikající úrovně zabezpečení i v horizontu desítek let pokroku v kryptoanalýze.

Šifra Serpent se skládá z těchto kroků:

- Počáteční permutace (*IP = Initial Permutation*), bez vlivu na kryptografické zabezpečení, pouze bitová úprava před vlastním šifrováním, vstupem je otevřený text (*P = plaintext*).
- 32 rund (*R = round*), kde každá runda spočívá v “namíchání” připraveného rundovního klíče k otevřenému textu (operace XOR), průchod substitučním boxem (*S*) a lineární transformací (*L*), která je ale vynechána v poslední rundě a nahrazena další operací XOR (připravených klíčů je tedy třeba mít o jeden více než rund – celkem 33). Mezivýsledek rundy je značen B_i, kde *i* značí pořadí rundy.
- Finální permutace (*FP = Final Permutation*), bez vlivu na kryptografické zabezpečení, pouze bitová úprava po vlastním šifrování, výstupem je kryptogram (*C = ciphertext*).

Šifrovací proces může být formálně popsán rovnicemi:

$$B_0 := IP(P),$$
$$B_{i+1} := R_i(B_i), pro\ i \in \{0;31\},$$
$$C := FP(B_{32}),$$

kde

$$R_i(X) = L(S_i(X \oplus K_i)), \text{pro } i \in \{0;30\},$$
$$R_i(X) = S_i(X \oplus K_i) \oplus K_{32}, \text{pro } i = 31.$$

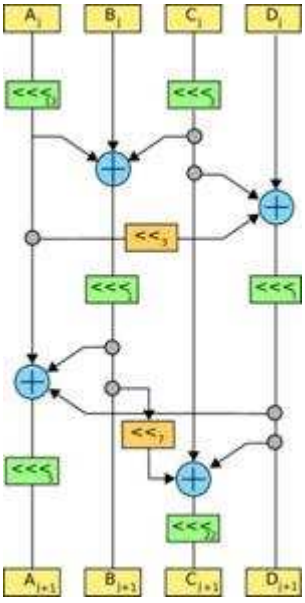
Grafické znázornění bez počáteční a finální permutace, které nemají vliv na kryptografické zabezpečení, je možné nalézt na obr. 2 v kapitole 3.1, obrázek obsahuje pro srovnání i šifrovací proces algoritmu AES (Rijndael).

2.2. Substituční boxy

Jeden substituční S-box pracuje pouze se čtyřmi bity, proto jich musí být paralelně celkem 32, aby bylo možné zpracovat zároveň 128 bitů. V šifrovacím systému se vyskytuje 8 odlišných typů těchto S-boxů, každý typ je používán ve čtyřech rundách a jak již bylo řečeno, každý je použit paralelně 32krát. Např. první typ S-boxu (S_0) je použit v rundách: R_0, R_8, R_{16} a R_{24}, druhý v rundách R_1, R_9, R_{17} a R_{25} atd. Více k popisu těchto substitučních boxů je možné nalézt např. v [6].

2.3 Lineární transformace

Lineární transformace probíhající v každé rundě algoritmu Serpent je znázorněna na obr. 1, převzato z [11]. Značka „<<< X“ má význam rotace bitů o uvedenou hodnotu X, značka „<< X“ značí posun bitů taktéž o uvedenou hodnotu. Značka $\oplus$ má význam logické operace XOR. V horní části obrázku je vstup, v dolní výstup transformace. Úkolem lineární transformace je maximalizace lavinového efektu, čímž se myslí to, že vliv změny jednoho bitu na vstupu ovlivní co největší počet bitů na výstupu.



Obr. 1: *Lineární transformace (nad čtyřmi 32-bitovými slovy) v jedné rundě při šifrování*

2.4 Schéma tvorby klíčů

Jak bylo uvedeno výše, při šifrování (a samozřejmě pak i při dešifrování) je potřeba celkem 33 rundovních klíčů, pokaždé ve formě čtyř 32-bitových slov, tedy celkem 132 těchto 32-bitových klíčů. Tyto klíče jsou získávány expanzí původního 256-bitového klíče, jehož vznik v případě kratšího uživatelského klíče (metoda výplně) byl popsán v úvodu kap. 2.1. Tento klíč je nejdříve rozdělen na osm 32-bitových slov (*words*) w_{-8}, …, w_{-1} a následně rozšířen pomocí následujícího rekurentního vztahu na tzv. předklíč (*prekey*) w_0, …, w_{131}.

$$w_i := (w_{i-8} \oplus w_{i-5} \oplus w_{i-3} \oplus w_{i-1} \oplus \phi \oplus i) \lll 11, \text{ pro } i \in \{0; 131\},$$

kde ϕ je hexadecimální hodnota `0x9e3779b9`.

Na takto získané předklíče jsou po čtveřicích následně aplikovány substituční S-boxy a výstupem jsou jednotlivé rundovní klíče (32 celkem) a jeden další klíč, který je použit po dokončení 32. cyklu.

2.5 Dešifrování

Dešifrovací proces probíhá odlišně, jsou použity inverzní S-boxy a inverzní lineární transformace v opačném pořadí, samozřejmě taktéž rundovní klíče jsou použity v převráceném pořadí, což je u dešifrování běžné. Struktura pro dešifrování tedy u tohoto algoritmu není úplně stejná jako pro šifrování, jelikož jak S-boxy, tak lineární transformace nejsou stejné jako u šifrování, ale navzájem inverzní.

3. Srovnání s AES (Rijndael)

3.1 Základní srovnání

Jak známo, Serpent se standardem AES nakonec nestal, tím je algoritmus Rijndael, proto je vhodné pokusit se tyto dva algoritmy srovnat. I protože obě tyto symetrické blokové šifry vznikly za stejným účelem a na základě stejného zadání, jejich hlavní principy jsou velmi podobné.

Rijndael stejně jako Serpent sestává z dvou hlavních oblastí – tvorba rundovních klíčů a vlastních rund. Počet rund u Rijndaelova algoritmu je poměrně nízký a variabilní podle délky klíče: 128-bitový má 10 rund, 192-bitový má 12 rund, 256-bitový má 14 rund, u Serpentu je počet rund nezávislý na délce šifrovacího klíče, ale zato poměrně vysoký, 32. U Rijndaelu je nejdříve provedeno počáteční kolo odlišné od průběhu dalších kol a taktéž poslední runda má jiný charakter (podobně jako u Serpentu).

U Rijndaelu jsou definovány celkem čtyři různé operace v rundách: *SubBytes* (substituce bajtů), *ShiftRows* (posuny bajtů v řádcích), *MixColumns* (násobení sloupců) a *AddRoundKey* (přičtení rundovního klíče – operace XOR). Všechny tyto čtyři operace jsou použity pouze ve standardní rundě, tj. ne v počáteční ani závěrečné. Schéma tvorby klíčů je v zásadě velmi podobné, u Serpentu jen o něco složitější. Rijndael tedy používá stejně jako Serpent jak substituční operace (zajištění konfúze – nelze určit závislost kryptogramu na šifrovacím klíči), tak permutační operace – rotace (difúze – nelze nalézt statistické závislosti kryptogramu na původním otevřeném textu) i lineární operace (zajištění konfúze i difúze). Grafické znázornění šifrovacího procesu je možné pro oba algoritmy nalézt na obr. 2.

Nejzásadnější rozdíl je tedy v počtu těchto rund a délce klíče. Je patrné, že čím je počet rund vyšší, tím by měla být konfúze i difúze větší. Serpent používá vždy 256-bitový klíč, délka u AES je možná ve třech krocích (128, 192, 256), přičemž nejčastější je použití nejkratší délky, tedy 128 bitů. Serpent by tedy měl poskytovat vyšší úroveň zabezpečení než Rijndael.

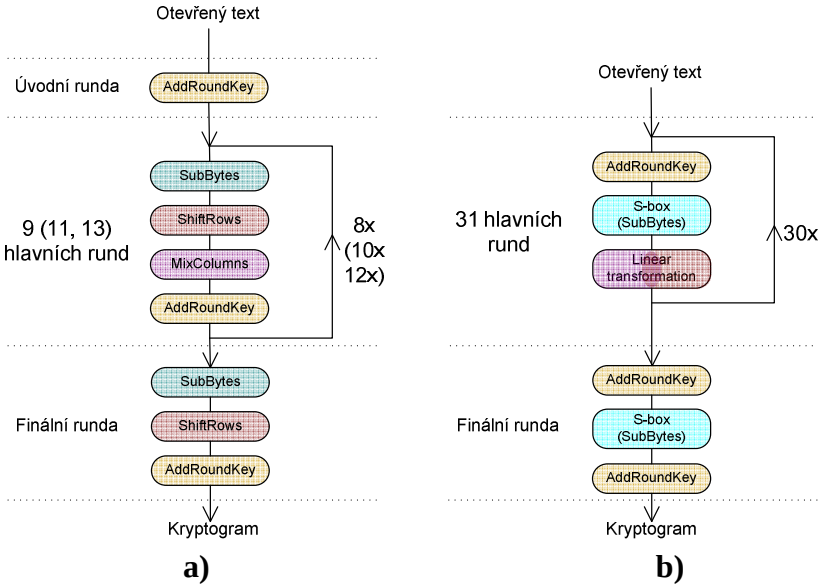
3.2 Výkonnostní a hardwarové srovnání

Jedním z nejvýznamnějších aspektů, který rozhodoval při výběru AES, byla kromě kvality také rychlost šifrování (a dešifrování) jednotlivých algoritmů. Srovnání výkonnosti softwarových implementací lze nalézt např. v [12]. Z těchto analýz vyplývá, že Rijndaelův algoritmus je ve všech činnostech poměrně dosti výrazně rychlejší než Serpent, což je ale zřejmé již ze samotného počtu prováděných rund.

Rijndael	Serpent
+ Malý počet kol (10 až 14)	– Velký počet kol (32)
+ Malý počet dílčích klíčů (11 až 15)	– Velký počet dílčích klíčů (33)
– Počet rund závisí na délce klíče	+ Pevný počet rund
+ Žádné složité matematické operace	+ Žádné složité matematické operace
+ Všechny rundy jsou identické	– 8 různých typů S-boxů
– S-boxy poměrně velké (8 x 8 bitů)	+ S-boxy menší (4 x 4 bity)
+ Velká část hw může být sdílena při šifrování a dešifrování	– Téměř žádný hw nemůže být sdílen při šifrování a dešifrování

Tab.1: *Srovnání šifer Rijndael a Serpent z pohledu hardwarové realizace*

V případě hardwarových implementací již je situace odlišná, Serpent je podle některých zdrojů rychlejší než Rijndael [13], přesto jsou ale mezi těmito algoritmy dost zásadní implementační rozdíly [14], viz Tab. 1.



Obr. 2: *Schéma šifrovacího procesu* **a)** *AES (Rijndael)* **b)** *Serpent*

5. Závěr

V článku byl popsán symetrický šifrovací algoritmus Serpent a také jeho srovnání se standardem AES – konkurentem ze soutěže o tento standard – šifrou Rijndael. Algoritmy mají mnoho podobného, významnější rozdíly byly vyzdviženy. Obě šifry jsou všeobecně považovány za dostatečně bezpečné, Serpent ale obsahuje větší rezervy do budoucnosti, což je logicky vykoupeno jeho nižší rychlostí. Významnou nevýhodou Serpentu je také nízká využitelnost šifrovací struktury pro dešifrování.

Jiné stanovisko je možné zaujmout, pokud se na tyto algoritmy podíváme z odlišného úhlu. AES (Rijndael), jakožto rozšířený standard, je neustále podrobován různým výzkumům a snahám o prolomení nebo nalezení významnější chyby, zatímco Serpent, jakožto méně známý algoritmus, zůstává mimo hlavní proud a mimo ohrožení. To může představovat významnou bezpečnostní výhodu, podobně jako je tomu mezi operačními systémy typu Windows a Linux.

Další stanovisko vychází z toho, že existují jisté dohady, zda není organizace typu americká vláda schopna již dnes AES luštit bez znalosti klíče, podobně jako se o tom spekulovalo u algoritmu DES. Serpent je díky svému robustnějšímu návrhu bezpečnější a tedy může představovat lepší řešení, pokud má uživatel pocit, že by mohl být sledován. Lze také využít kombinace těchto algoritmů, což v praxi vypadá tak, že se data šifrují dvakrát, jednou pomocí jednoho algoritmu a následně kryptogram zašifrujeme znovu pomocí druhého algoritmu, pokaždé s jiným klíčem. Tím je dosažena vysoká odolnost proti potenciálním kryptoanalytickým útokům na bezpečnost šifry, nesmí se však zapomínat na ostatní typy útoků, jako jsou sociální inženýrství, postranní kanály aj.

Literatura

[1] Internetová stránka americké vládní organizace NIST (National *Institut of Standards and Technology*). Dostupné na: http://www.nist.gov/.

[2] Data Encryption Standard, Wikipedia – volná encyklopedie (online). Dostupné na: http://en.wikipedia.org/wiki/Data_Encryption_Standard.

[3] IBM Corporation: MARS – a candidate cipher for AES (online), 1999. Dostupné na: http://www.research.ibm.com/security/mars.pdf.

[4] RSA Laboratories: The RC6 block cipher (online), 1998. Dostupné na: ftp://ftp.rsasecurity.com/pub/rsalabs/rc6/rc6v11.pdf.

[5] Daemon, J., Rijmen, V.: AES Proposal: Rijndael (online), 1999. Dostupné na: http://csrc.nist.gov/archive/aes/rijndael/Rijndael-ammended.pdf.

[6] Anderson, R., Biham, E., Knudsen, L.: Serpent – A candidate block cipher for the Advanced Encryption Standard (online), 1998. Dostupné na: http://www.cl.cam.ac.uk/~rja14/Papers/serpent.pdf.

[7] Schneier, B.: Twofish - a 128-bit block cipher (online), 1998. Dostupné na http://www.schneier.com/paper-twofish-paper.pdf.

[8] Bernstein, D. J.: Cache-timing attacks on AES (online), 2005. Dostupné na: http://cr.yp.to/antiforgery/cachetiming-20050414.pdf

[9] Stránky volného šifrovacího programu Truecrypt, dostupné na: http://www.truecrypt.org/docs/?s=cascades

[10] Schneier, B. a kol.: The Twofish Team´s Final Comments on AES Selection (online), 2000. Dostupné na: http://www.schneier.com/paper-twofish-final.pdf.

[11] Serpent (cipher), Wikipedia – volná encyklopedie (online), Dostupné na: http://en.wikipedia.org/wiki/Serpent_%28cipher%29.

[12] Schneier, B., Whiting, D.: A Performance Comparison of the Five AES Finalists (online), 2000. Dostupné na: http://www.schneier.com/paper-aes-comparison.pdf.

[13] Anderson, R., Biham, E., Knudsen, L.: The Case for Serpent (online), 2000. Dostupné na: http://www.cl.cam.ac.uk/~rja14/Papers/serpentcase.pdf.

[14] Lutz, A. K. a kol.: 2Gbit/s Hardware Realizations of Rijndael and Serpent:
A Comparative Analysis (online), 2003. Dostupné na:
http://www.springerlink.com/content/mx19nj1dum1cuwqu/fulltext.pdf

D. O čem jsme psali v červnu 2000 – 2007

Crypto-World 6/2000

Příloha : Navajo Code Talkers, revize z 15.6.1945, soubor Dictionary.htm

Crypto-World 6/2001

Příloha : priloha6.zip
(fotografie Security 2001, témata přednášek na konferenci Eurocrypt´2001)

Crypto-World 6/2002

Crypto-World 6/2003

Crypto-World 6/2004

Crypto-World 6/2005

Crypto-World 6/2006

Crypto-World 6/2007

E. Závěrečné informace

1. Sešit

Crypto-World je oficiální informační sešit "Kryptologické sekce Jednoty českých matematiků a fyziků" (GCUCMP). Obsahuje články podepsané autory. Případné chyby a nepřesnosti jsou dílem P.Vondrušky a autorů jednotlivých podepsaných článků, GCUCMP za ně nemá odbornou ani jinou zodpovědnost.
Adresa URL, na níž můžete najít tento sešit (zpravidla 3 týdny po jeho rozeslání) a předchozí sešity GCUCMP, denně aktualizované novinky z kryptologie a informační bezpečnosti, normy, standardy, stránky některých členů a další související materiály:
http://crypto-world.info

2. Registrace / zrušení registrace

Zájemci o e-zin se mohou zaregistrovat pomocí e-mailu na adrese pavel.vondruska@crypto-world.info (předmět: Crypto-World) nebo použít k odeslání žádosti o registraci elektronický formulář na http://crypto-world.info. Při registraci vyžadujeme pouze jméno a příjmení, titul, pracoviště (není podmínkou) a e-mail adresu určenou k zasílání kódů ke stažení sešitu.
Ke zrušení registrace stačí zaslat krátkou zprávu na e-mail pavel.vondruska@crypto-world.info (předmět: ruším odběr Crypto-Worldu!) nebo opět použít formulář na http://crypto-world.info. Ve zprávě prosím uveďte jméno a příjmení a e-mail adresu, na kterou byly kódy zasílány.

3. Redakce

E-zin Crypto-World

Redakční práce:	Pavel Vondruška
Stálí přispěvatelé:	Pavel Vondruška
	Jaroslav Pinkava
Jazyková úprava:	Jakub Vrána
Přehled autorů:	http://crypto-world.info/obsah/autori.pdf
NEWS	Vlastimil Klíma
(výběr příspěvků,	Jaroslav Pinkava
komentáře a	Tomáš Rosa
vkládání na web)	Pavel Vondruška
Webmaster	Pavel Vondruška, jr.

4. Spojení (abecedně)

redakce e-zinu	ezin@crypto-world.info ,	http://crypto-world.info
Vlastimil Klíma	v.klima@volny.cz ,	http://cryptography.hyperlink.cz/
Jaroslav Pinkava	Jaroslav.Pinkava@zoner.cz ,	http://crypto-world.info/pinkava/
Tomáš Rosa	t_rosa@volny.cz ,	http://crypto.hyperlink.cz/
Pavel Vondruška	pavel.vondruska@crypto-world.info ,	http://crypto-world.info/vondruska/index.php
Pavel Vondruška,jr.	pavel@crypto-world.info ,	http://webdesign.crypto-world.info
Jakub Vrána	jakub@vrana.cz ,	http://www.vrana.cz/