Kryptologie pro praxi – lidová tvořivost se nevyplácí

Hlavní architekt projektu GnuPG (též GPG) [2] se 27. listopadu minulého roku ocitl ve zvlášť nepříjemné situaci. Musel vydat zprávu [3], jejíž text byl jasný, stručný a mrazivý: *Byla nalezena závažná slabina v implementaci ElGamal... Toto je reálná, celosvětová slabina, která umožňuje získat během několik sekund váš privátní klíč...*

Všechny podepisovací klíče schématu ElGamal používané v systému GPG verze 1.0.2 (leden 2000) až 1.2.3 (srpen 2003) musí být považovány za kompromitované... Lze si domyslet, že člověku, který podstatnou část svého času věnuje právě tomuto volně dostupnému nástupci PGP [5] (o kterém dodnes není jasné, jak vážně to s ním vývojáři míní), se něco takového určitě nepsalo s lehkou rukou. V nelehké situaci jsme i my, neboť nechceme v žádném případě jinak slibnému projektu GPG ublížit, avšak zároveň cítíme, že toto je přesně ten druh poučení, které bychom si v našem seriálu měli rozebrat. Pokuste se proto v dalším textu ignorovat to, že se zrovna jedná o GPG – stejně tak dobře by zde totiž mohla stát jména řady jiných, podstatně honosnějších projektů. GPG samotné se z této rány naštěstí oklepalo a nyní z ní už může jen těžit.

Jak jsme už naznačili, řeč bude o chybě v implementaci podpisového schématu z rodiny ElGamal (ST 6/2004), která byla v dotčeném produktu skryta téměř 4 roky a která způsobila, že na základě znalosti jediné podepsané zprávy bylo možné na běžném PC za méně než sekundu vypočítat hodnotu privátního klíče [4]. Snad ani není nutné dodávat, že pokud by se něco takového stalo v době, kdy bude elektronické podepisování dokumentů rutinní záležitostí, jednalo by se o katastrofu překračující vize hollywoodských trháků. Drobnou náplastí sice je, že ElGamal nebyl v GPG nikdy preferovanou volbou (tou bylo DSA, případně RSA – ST 4/2004 a ST 3/2004), avšak zákon schválnosti funguje perfektně, takže v praxi by se bylo našlo určitě několik důležitých businessmanů, kteří by byli toto schéma používali. Proto si případ zasluhuje pozornost, i když jde jen o „nějaký" ElGamal.

Popis implementace

V ST 6/2004 jsme si podpisové transformace schématu ElGamal neuváděli, takže si je zde doplníme a to konkrétně v podobě použité v [2]: Veřejnými parametry instance je dvojice (p, g), kde p je prvočíslo a g je generátor $\mathbf{Z}_p^*$. Privátní podepisovací klíč x je celé číslo z intervalu $(0, p-1)$ a k němu příslušející veřejný klíč y je počítán jako $y=g^x \bmod p$. Podpisem zprávy M je dvojice celých čísel (r, s) vypočtených tímto postupem: Nejprve je vypočten hašový kód zprávy $h(M)$, který je formátován podle standardu EMSA-PKCS1-v1_5 (ST 10/2003). Výsledkem formátování je celé kladné číslo m, $m<p-1$. Dále program vygeneruje klíč zprávy (pojem viz ST 4/2004) k, $\gcd(k, p-1)= 1$, a počítá $r=g^k \bmod p$, $s=(m-xr)k^{-1} \bmod (p-1)$, kde $kk^{-1}\equiv 1 \pmod{p-1}$. Ověření podpisu (r, s) zprávy M probíhá následovně: Nejprve se ověří, že $0<r<p$. Poté je opět hašováním a formátováním vypočtena hodnota m, se kterou se ověří, že $y^r r^s \equiv g^m \pmod p$. Pokud tato kongruence platí (tj. obě strany mají stejný zbytek po dělení prvočíslem p), je podpis uznán jako platný. Dosazením snadno ověříme, že pro korektně vygenerovaný podpis platí: $y^r r^s \equiv g^{xr+ks} \equiv g^{xr+m-xr} \equiv g^m \pmod p$.

Kde byla slabina

Jednou z nevýhod ElGamalu je, že pracuje s poměrně velkými exponenty, což se u slabších platforem může zřetelně projevovat na výkonu aplikace. Proto autoři přišli se spásným nápadem: Místo aby čísla x a k (viz výše) vybírali náhodně z celého intervalu $(0, p-1)$, jak to vyžaduje originální návrh, rozhodli se, že bude stačit vybírat je tak, aby jejich bitová délka odpovídala hodnotě $3q_{\text{bit}}/2$, kde q_{bit} je hodnota závisející podle takzvané Wienerovy tabulky na délce l čísla p. Například pro typickou délku p 1024 bitů je $q_{\text{bit}}=165$. Přitom pro všechna přípustná l platí, že $4q_{\text{bit}}<l$. Navíc u hodnoty k není požadováno, aby byl nejvyšší bit nastaven (to už je zde ale nevýznamný detail). Na první pohled by se mohlo zdát, že $3*165/2\cong 248$ bitů entropie je dost na to, aby útočník nemohl schéma prolomit, avšak chyba lávky! Ve složitosti úloh luštění zejména asymetrických schémat totiž hraje podstatnou roli nejen „absolutní" velikost bezpečnostních parametrů, ale i jejich struktura a vzájemné vztahy. Díky tomu, že velikosti x a k jsou méně než poloviční v porovnání s modulem p, tak je zde 248 bitů proklatě málo, i kdyby se nám to jinak zdálo dost.

Ačkoliv práce [4] útok neuvažuje jako útok postranním kanálem, lze ho na něj převést. Pak lze tvrdit, že popsaná modifikace vlastně mimoděk vytváří podprahový a nad ním postavený kleptografický kanál (ST 3/2003), který s každým podpisem vyzařuje informaci o privátním klíči. Díky velmi malé poměrné velikosti k má tento kanál tak vysokou kapacitu, že se hodnota x (rovněž malá) celá pohodlně vyzáří jedním jediným podpisem. Heuristickou souvislost mezi entropií k a kapacitou vzniklého podprahového kanálu si lze dobře uvědomit z [1] (tam popsaný kanál ovšem není zcela totožný s naším).

Zjednodušený popis samotného útoku je následující: Útočník nejprve běžným způsobem získá veřejné parametry, veřejný klíč a jednu podepsanou zprávu (stačí její zformátovaná podoba m). Z definice podepisovací transformace platí, že

$$sk+rx\equiv m \pmod{p-1}. \qquad (1)$$

Při správné implementaci se tento triviálně odvozený vztah pochopitelně nedá k napadení využít, avšak útočník ví, že obě hodnoty k a x jsou velmi malé, a tak se snaží dál. Jak je ukázáno v [4], tvoří množina všech celočíselných dvojic (u, v) řešících kongruenci

$$su+rv\equiv 0 \pmod{p-1} \qquad (2)$$

dvourozměrnou číselnou mřížku L, $L\subset\mathbf{Z}^2$, s determinantem $\det(L)=(p-1)/e$, kde $e=\gcd(r, s, p-1)$. Dále je ukázáno, že umíme snadno najít bázi této mřížky. Útočník nyní pomocí rozšířeného Euklidova algo-

ritmu nalezne libovolnou dvojici (k', x') řešící kongruenci (1). Všimněme si, že vektor $\boldsymbol{w}=(k'-k, x'-x)$ patří do L, neboť je řešením (2). Až sem se lze dostat i za normálních okolností a stále to nic neznamená, neboť zůstává neschůdným najít tu „pravou" dvojici (k, x), respektive ten pravý vektor $\boldsymbol{w}$. Díky spásnému nápadu architektů tu však nekončíme, ale začínáme. S ohledem na známou malou velikost k a x víme, že vektor $\boldsymbol{w}$ je relativně blízko vektoru $\boldsymbol{t}=(k'-2^{3q_{\text{bit}}/2-1}, x'-2^{3q_{\text{bit}}/2-1})$, jejichž vzdálenost je v Euklidově normě řádově blízká hodnotě $2^{(3q_{\text{bit}}-1)/2}$, přičemž je podstatné, že tato vzdálenost je výrazně menší než hodnota $\det(L)^{1/2}$. S ohledem na běžnou heuristiku používanou v kryptoanalýze tak můžeme předpokládat, že vektor $\boldsymbol{w}$ je z celé mřížky L k vektoru $\boldsymbol{t}$ nejblíže, takže jej lze najít řešením problému CVP (Closest Vector Problem), což je pro tento typ mřížky triviálně schůdná úloha. Jakmile vektor $\boldsymbol{w}$ nalezneme, jsme u cíle – jeho odečtením od vektoru (k', x') získáme na druhé souřadnici výsledku přímo hledanou hodnotu privátního klíče x. Doplňme, že algoritmus útoku je sice pravděpodobnostní, avšak pravděpodobnost neúspěchu je prakticky zanedbatelná.

Poučení

Začneme tím základním a to náhodnými čísly v kryptografii: V současných schématech se s nimi setkáme prakticky všude a v drtivé většině případů na jejich kvalitě silně závisí bezpečnost celého schématu. Přesto jsme si mohli všimnout, že pojem náhodné číslo chápe řada architektů jako obtížný hmyz, který poletuje kolem implementovaného algoritmu a kterého je záhodno se co nejjednodušším způsobem rychle zbavit, aby „to šifrování" začalo konečně fungovat a oni se mohli věnovat „tomu důležitému". V důsledku toho se nám potom čísla svévolně krátí, deformuje se jejich rozdělení nebo se dokonce rovnou opakují. K vidění jsou i případy, kdy se aplikace odladí s běžným (a z bezpečnostního hlediska zcela nepřijatelným) generátorem ze standardních knihoven Céčka či Pascalu, načež se toto dočasné řešení stane zakrátko řešením trvalým. Popsaný útok je přitom jen jedním z mnoha důsledků, kam může nevhodné zacházení s náhodnými čísly zajít. Berme ho proto jako důrazné varování před všemi takovými praktikami.

Zobecněným poučením je, že aplikační architekti by se rozhodně neměli v implementovaných schématech pouštět do žádných „kryptologických lidových tvořivostí" a už vůbec ne do takových, které viditelně uvolňují sílu některého z bezpečnostních parametrů. To už nemluvíme jen o náhodných číslech. Výstražné světlo zde ovšem svítí i pro kryptology: Záměrně jsme si v popisu útoku dobře označili místo, odkud se teprve začíná diskutovaná slabina projevovat. Snad je z toho dobře patrné, že její využití není (kromě jistých náznaků heuristického rázu) vidět hned od začátku, ale že je třeba konkrétní druh útoku do detailů promyslet, aby bylo možné říci, jestli a nakolik se zamýšlená úprava projeví jako slabina. Uvážíme-li ohromné množství teoretických útoků, zjistíme, že prověření vlivu takové modifikace rozhodně není triviální záležitost. Dalším poučením proto je, že lépe než ořezávat stávající mechanizmus, je vhodnější jej rovnou celý opustit a zkusit najít nějakého vhodnějšího, prověřeného kandidáta. Teprve, když tato možnost selže, je načase „řezat". Ovšem vždy s citem...

Vlastimil Klíma, Tomáš Rosa,
v.klima@volny.cz, trosa@ebanka.cz

LITERATURA

[1] Anderson, R., Vaudenay, S., Preneel, B., Nyberg, K.: The Newton Channel, in Proc. of Information Hiding '96, 39–48, Springer-Verlag, 1996

[2] GPG – The GNU Privacy Guard, http://www.gnupgp.org

[3] Koch, W.: GnuPG's ElGamal signing keys compromised, veřejné internetové oznámení, 27. listopadu 2003

[4] Nguyen, P.-Q.: Can We Trust Cryptographic Software? Cryptographic Flaws in GNU Privacy Guard v1.2.3, Proc. of Eurocrypt '04, 151–176, Springer-Verlag, 2004

[5] PGP – Pretty Good Privacy, http://www.pgp.com

[6] E-archivy http://cryptography.hyperlink.cz a http://crypto.hyperlink.cz

Zákon o elektronických komunikacích

Vláda na svém prvním zářijovém zasedání schválila na návrh ministra informatiky V. Mlynáře dlouho diskutovaný text zákona o elekronických komunikacích, který po projednání v Poslanecké sněmovně a poté i v Senátu ČR nahradí telekomunikační zákon z roku 2000, a do českého právního řádu trvale zakotví nový regulační rámec sjednocující pravidla chování účastníků telekomunikačního trhu ve všech zemích EU. Cílem zákonné normy je rozšířit a jednoznačně ohraničit předmět regulace, zjednodušit některé rozhodovací postupy a podpořit zdravý vývoj konkurenčního prostředí. Příkladem působení na zvyšování transparentnosti telekomunikačního trhu a udržení rovného postavení podnikatelských subjektů je ustanovení o přenositelnosti čísla i v sítích mobilních operátorů (ČTc už službu přenositelnosti čísel ve své síti zavedl). Zákon dále přesně vymezuje obecná pravidla pro řešení sporů vznikajících v souvislosti s poskytováním telekomunikačních služeb v liberalizovaném evropském prostředí, mění organizační strukturu ČTÚ, vymezuje úlohu ČTÚ při řešení konfliktů a definuje přesná pravidla při udělování sankcí.

Některé rozhodovací pravomoce Rady pro rozhlasové a televizní vysílání (zejména ty, které určují technické podmínky provozu, např. přidělování frekvencí v souvislosti se zaváděním digitálního vysílání) budou v budoucnu převedeny na Český telekomunikační úřad (ČTÚ). Ten bude řídit pětičlenná rada odborníků jmenovaná vládou na dobu pět let v čele s předsedou. Už nyní se objevují obavy, že za plat vysokého státního úředníka se nenajde potřebný počet kandidátů s odpovídající erudicí a kvalifikací.

Zákon dále umožní zavést tzv. asymetrickou regulaci, to znamená, že budou přísněji posuzovány strategické záměry i prohřešky dominantních operátorů s „významnou tržní silou" ve srovnání s menšími a začínajícími firmami, které mohou za určitých podmínek počítat v souladu s relevantními směrnicemi EU s definovanou ochranou a podporou. Cílem je usnadnit novým společnostem vstup na trh a vytvořit tak vhodné prostředí pro vznik a rozvoj nových sítí a služeb. Zákon rovněž ruší dosavadní licenční systém a obecně zjednodušuje pravidla pro podnikání v oblasti elektronických komunikací. Výkon činnosti už nebude podléhat živnostenskému zákonu, ale pouze registraci u ČTÚ.

Koncepce univerzální služby (kam patří mimo jiné například i provozování a udržování sítě veřejných telefonních automatů), to znamená obsluha vymezeného souboru funkcí, které musí být ve veřejném zájmu ve stanovené kvalitě dostupné na celém území státu (minimálně v rozsahu, který v současnosti nabízí ČTc) všem občanům za všeobecně přijatelnou cenu, zůstává zachována. Každou dílčí službu však může provozovat jiná společnost, aniž by nutně pokryla celé území státu. Pověření poskytovatelů univerzální služby bude podmíněno výběrovým řízením, za jehož vypsání a průběh bude odpovědný regulátor. Ztráta z provozování univerzální služby bude kryta z účtu, do něhož budou přispívat podle poměrného podílu na trhu všechny podnikatelské subjekty rezortu. Hospodaření s účtem musí být veřejné. Zvýhodnění sociálně slabých a těžce zdravotně postižených občanů zůstane zachováno přibližně na současné úrovni.