Bezpečnostní kódy, díl 8.

V klidu a bezpečí (8)

Až dosud jsme se zabývali zejména bezpečnostními kódy lineárního typu. Počínaje tímto pokračováním se postupně přesuneme do teorie kódů cyklických, které tvoří důležitou a v praxi často používanou skupinu ECC.

Ještě než se pustíme do slíbeného tématu, dovolím si provést malou odbočku a vysvětlit, jakým směrem se bude styl výkladu v několika příštích článcích ubírat a proč. Vycházím z toho, že tento seriál má sloužit hlavně k pochopení teorie kolem ECC a k ukázání, že věci nemusí být tak magické a nesrozumitelné – pouze je třeba se zaobírat i takovými "detaily", které se obvykle považují za "zbytečné pitvání" tématu, a tudíž se jaksi neříkají.

Na druhou stranu by bylo příliš naivní myslet si, že na takto malém prostoru je možné prezentovat celou teorii ECC a navíc ještě formálním způsobem. Budeme se zde proto snažit poukázat zejména na hlavní principy a souvislosti, o které se teorie bezpečnostních kódů opírá. Nepůjde nám přitom ani o podání zcela přesného formálního popisu, ani o vytvoření monografie, se kterou si jako s jediným zdrojem informací vystačíme při implementaci ECC. Naším cílem bude si vždy příjemně odpočinout a popřemýšlet nad velmi zajímavou matematickou teorií, která se může navíc chlubit bohatým praktickým uplatněním.

Pro úplnost ještě připomínám, abyste neváhali použít mou e-mailovou adresu, kdykoliv budete mít jakékoliv připomínky či dotazy k probíranému tématu.

Algebraické struktury

Až dosud jsme se v probíraných tématech mohli spokojit s tím, že jsme používali běžné matematické operace "obvyklým" způsobem a příliš jsme nepátrali po tom, jak moc bylo naše počínání korektní. Budeme-li však chtít správně pochopit základy cyklických kódů, nezbude nám, než přestat se spoléhat na ony obvyklé principy a říci si pár slov o základních algebraických strukturách a o způsobu jejich používání.

Obecně budeme za algebraickou strukturu považovat nějakou množinu hodnot M, na které je definována jedna nebo více operací, které jsou na této množinou uzavřené (tj. pokud vstupní hodnoty příslušné operace patří do M, potom je i výsledek této operace prvkem množiny M).

Konkrétně se zatím omezíme na binární operace, což jsou zobrazení f: MxM → M. Snadno určíme, že všech takových zobrazení (tj. binárních operací na M) je $|M|^{|M|^2}$. Většina z nich však není pro další teoretické studium příliš přínosná, takže při zavádění nových operací se většinou vychází z jemných modifikací známých operací "+" a "*". Obvykle jim ponecháváme i jejich původní název, tj. operace sčítání a násobení.

Je však třeba mít na zřeteli, že konkrétní výpočet uvedených operací může silně záviset na konkrétní množině M, na které jsou definovány. Celkem snadno se můžeme v teorii setkat s operací, které se sice říká násobení, ale která má ke známému násobení na tělese reálných čísel velmi daleko. Co se naopak u těchto operací nemění, jsou jejich vlastnosti, podle kterých je možné provádět klasifikaci.

Na obrázku 1 je uvedena tabulka algebraických struktur, se kterými se budeme v teorii ECC setkávat nejčastěji. Zde uvedené rozdělení předpokládá, že máme množinu M, na které jsme definovali jednu nebo dvě binární operace, které značíme symboly "+" a "*". Pokud tyto operace splňují podmínky uvedené v levém sloupci tabulky, potom příslušnou dvojici (M, op_1) nebo trojici (M, op_1, op_2) označujeme názvem, který je uveden v pravém sloupci tabulky.

Z obrázku například vidíme, že množinu, na které je definována operace sčítání s příslušnými vlastnostmi, označujeme jako aditivní grupu, analogicky množinu s definovanou operací násobení jako grupu komutativní. Grupy pro nás budou představovat základní stavební prvek složitějších struktur, jako jsou okruhy a tělesa. Vzhledem k názvům uvedeným na obrázku 1 poznamenejme, že označení "komutativní okruh se jednotkovým prvkem" budeme zkracovat na termín "okruh". To můžeme udělat, protože s jiným typem okruhů zde prozatím nebudeme pracovat.

Obě struktury – těleso i okruh – se vyznačují tím, že mají definovánu jak operaci sčítání, tak násobení. Rozdíl mezi tělesem a okruhem je v tom, že v okruhu na rozdíl od tělesa existují prvky, které

vzhledem k operaci násobení nemají v M inverzní prvek. Zatímco tedy těleso můžeme považovat zároveň za aditivní a multiplikativní grupu, okruh je pouze grupou aditivní. Operace násobení zde sice existuje také, avšak netvoří grupu.

Příkladem struktury, která je pouze okruhem, může být například okruh celých čísel (Z). Tato struktura je sice aditivní grupou (ke každému číslu x existuje jeho aditivní inverze -x), ale není grupou multiplikativní (s výjimkou prvku 1 neobsahuje Z pro žádný prvek x také prvek 1/x). Tělesem je teprve množina racionálních čísel, která na rozdíl od Z obsahuje ony “chybějící" zlomky. Poznamenejme, že tělesem je také množina reálných čísel, avšak zde se jedná o zcela odlišný druh struktury, než s jakou se budeme setkávat. Těleso reálných čísel je totiž spojité a nekonečné, zatímco námi studované struktury budou diskrétní a konečné.

Věnujme se v krátkosti pojmu konečné těleso. S přívlastkem “konečný" se budeme v teorii ECC setkávat velmi často a můžeme jej použít pro každou výše uvedenou algebraickou strukturu. Význam tohoto přívlastku snad ani nemá cenu nějak formalizovat, neboť plně odpovídá jeho intuitivnímu chápání – daná struktura (množina M) má konečně mnoho prvků. Konečná tělesa se většinou označují jako Galoisova tělesa a značí se GF(q), kde q udává počet prvků v tomto tělese.

Při studiu literatury se můžete setkat s nejrůznějšími definicemi tělesa GF(q) (nejčastěji jako rozšíření nějakého konečného tělesa F s charakteristikou p – viz [ADAM89]), avšak námi zavedená definice je pro nás zatím nejen postačující, ale díky tomu, že každé konečné těleso je izomorfní s nějakým Galoisovým tělesem (důkaz viz [ADAM89]), i korektní.

V souvislosti s konečnými tělesy byla dokázána následující věta: Pro každé konečné těleso GF(q) platí, že $q = p^n$, kde p je prvočíslo a $n \in N\backslash\{0\}$ – tvrzení T8.1. Důsledkem tohoto tvrzení je, že existují pouze taková konečná tělesa, která mají počet prvků rovný mocnině nějakého prvočísla. Odtud například plyne, že nemůže existovat těleso GF(6) (tělesa GF(4) a GF(16) naproti tomu existují).

Jistě je nyní zajímavé se ptát, jestli je možné implikaci v tvrzení T8.1 obrátit, nebo jestli naopak existují i takové mocniny prvočísel, pro které GF(q) tělesem není. Ukazuje se, že T8.1 obrátit lze, díky čemuž dostáváme následující tvrzení: Pro každé prvočíslo p a každé celé kladné číslo n existuje konečné těleso GF(q), $q = p^n$ – tvrzení T8.2. Důkaz uvádí například [VAOO89] a [ADAM89].

Na závěr této části poznamenejme, že ačkoliv jsme se zde věnovali nejvíce problematice konečných těles, v teorii ECC si velmi často vystačíme i se strukturou, kterou jsme zde nazvali okruh. Jak už víme, má okruh oproti tělesu jedinou nevýhodu, že není zaručena existence inverzního prvku pro operaci násobení. Pokud ovšem tuto vlastnost nepožadujeme, může být užití okruhu naopak výhodnější, neboť (jak uvidíme později) nejsme například při konstrukci rozšíření nějakého tělesa F pomocí zbytkových tříd polynomu f(x) nuceni volit pouze ta f(x), která jsou nad F ireducibilní.

Polynomy nad tělesem F

Pro další výklad budeme předpokládat, že máme dáno nějaké konečné těleso F. Naším cílem bude nad tímto tělesem vybudovat nějakou další algebraickou strukturu, která bude mít rovněž vlastnosti tělesa či okruhu. Tomuto postupu se obecně říká rozšíření tělesa F a pro první přiblížení si můžeme uvést analogii s vektorovým prostorem, který je v podstatě také rozšířením nějakého tělesa (v našem případě opět konečného).

Začneme opět příznačně, a to definicí pojmu polynom: Polynomem nad tělesem F rozumíme výraz $a(x) = a_0 + a_1x +...+ a_nx^n$, kde $a_i \in F$, $0 \leq i \leq n$ a koeficient a_0 označujeme jako konstantní člen – definice D8.1. Poznamenejme, že v teorii ECC se polynomy obvykle zapisují od nejnižší mocniny po nejvyšší, což je dáno snahou o přizpůsobení se systému číslování souřadnic v aritmetických vektorech, což následně umožňuje snadné mapování vektorů na koeficienty polynomů a naopak.

Důležitým parametrem polynomu a(x) je jeho stupeň, který značíme deg(a(x)) a definujeme jako nejvyšší číslo k, pro které platí $a_k \neq 0$, stupeň nulového polynomu přitom definujeme deg(a(x)) = -1. Polynom a(x), pro který platí deg(a(x)) = 0, nazýváme konstantní polynom – definice D8.2. Polynom a(x), pro který platí $a_{deg(a(x))} = 1$, nazýváme normovaný – definice D8.3.

Vezměme si nyní množinu všech polynomů nad tělesem F a označme ji jako F[x]. Naším cílem bude nyní ukázat, že tato množina spolu s operacemi sčítání a násobení polynomů tvoří okruh. Začneme definicí operace sčítání: Mějme polynomy $a(x), b(x) \in F[x]$. Pro polynom $c(x) = a(x) + b(x) = c_0 + c_1x +...+ c_nx^n$ potom platí: $c_i = a_i + b_i$ – definice D8.3. Připomeňme, že pro účely sčítání koeficientů zde používáme operaci sčítání tak, jak je definována na příslušném tělese F (tj. může to být například součet celých čísel modulo p – pokud $F = Z_p$, atd.).

Vidíme, že definice sčítání na F[x] je v podstatě velmi intuitivní záležitostí. Obdobně je tomu i v případě násobení polynomů a(x), b(x) ∈ F[x], kde pro koeficienty polynomu c(x) = a(x)*b(x) platí: $c_i = a_0b_i + a_1b_{i-1} + ...+ a_ib_0$ – definice D8.4. Sčítání a násobení koeficientů se zde opět provádí podle pravidel definovaných pro příslušné těleso F.

Při definici operací sčítání a násobení na F[x] jsme zároveň ukázali i jejich uzavřenost (součet i součin dvou polynomů z F[x] je opět polynomem z F[x]). Ověřit zbývající podmínky a přesvědčit se tak, že F[x] je opravdu okruh, je již víceméně jen rutinní záležitostí.

V případě operace součinu dvou polynomů na F[x] můžeme dokázat následující pomocná tvrzení: Pro dva nenulové polynomy a(x), b(x) platí, že deg(c(x)=a(x)*b(x)) = deg(a(x)) + deg(b(x)) – tvrzení T8.3. Důsledkem tohoto tvrzení je, že pokud platí a(x)*b(x) = 0, potom je alespoň jeden z polynomů a(x), b(x) nulový – tvrzení T8.4. Důkaz je snadný, neboť pokud by platilo a(x)*b(x) = 0 pro nějaké nenulové polynomy a(x) a b(x), potom by platilo, že deg(a(x)*b(x)) = -1, což je spor s T8.3.

Dalším důsledkem tvrzení T8.3 také je, že F[x] není těleso – tvrzení T8.5. Předpokládejme nějaký polynom a(x) ∈ F[x], kde deg(a(x)) > 0. Pokud by k tomuto polynomu existovala multiplikativní inverze, tj. nějaké nenulové $a(x)^{-1}$ takové, že $a(x)^* a(x)^{-1} = 1$, potom by platilo, že $\deg(a(x)^* a(x)^{-1}) = \deg(1) = 0$, a to je opět spor s T8.3.

Dělení polynomů

Okruh F[x], který jsme si právě zavedli, má vzhledem k našemu záměru studovat teorii ECC podstatnou nevýhodu: není konečný. Naše další snažení proto bude směřovat k vytvoření “obdobné" struktury, která však již bude konečná.

Abychom mohli zamýšlenou úpravu provést, musíme si nejprve definovat operaci dělení polynomů. Uveďme si nejprve užitečné tvrzení: Pro libovolné polynomy a(x), b(x) ∈ F[x], b(x) ≠ 0, existuje právě jedna dvojice polynomů q(x), r(x) ∈ F[x], taková, že a(x) = q(x)*b(x) + r(x), kde deg(r(x)) < deg(b(x)) – tvrzení T8.6. Obdobně jako v případě celých čísel nazýváme polynom q(x) podílem a polynom r(x) zbytkem po dělení.

Základní algoritmus pro dělení polynomů na F[x] silně připomíná běžný postup dělení celých čísel. Pro lepší ilustrativnost si jej uvedeme jako příklad na obrázku 2. Zde je vyobrazen způsob dělení dvou polynomů a(x), b(x), pokaždé nad třemi různými tělesy. Vidíme, že vlastní postup je jednoduchý a spočívá v určování koeficientů podílu na základě podílu koeficientů u nejvyšších mocnin polynomů a(x) a b(x). Poté provedeme odečtení odpovídajícího násobku polynomu b(x) od a(x) a se získaným výsledkem a(x)' pokračujeme rekurzivně v určování zbývajících koeficientů polynomu q(x). Jakmile v průběhu dělení obdržíme polynom a(x)', deg(a(x)') < b(x), položíme r(x) = a(x)' a proces dělení ukončíme.

Záměrně jsme si uvedli výsledky dělení syntakticky stejných polynomů nad třemi různými tělesy, abychom si ilustrovali, jak základní operace na F ovlivňují operace na F[x]. Zajímavým námětem pro zamyšlení může být fakt, že koeficienty obdržených polynomů jsou sice v tělese Z různé, avšak v příslušných Z_p náleží vždy ke stejným třídám ekvivalence, čili jsou spolu kongruentní. Poznamenejme také, že zatímco nad Z_3 je polynom a(x) dělitelný polynomem b(x), nad tělesy Z a Z_2 tomu tak není.

Již jsme se zmínili o pojmu ireducibilní polynom, takže nyní si uvedeme jeho definici: Polynom f(x) je ireducibilní nad tělesem F, pokud jej není možné vyjádřit součinem f(x) = a(x)*b(x), kde a(x), b(x) jsou polynomy okruhu F[x] nižšího stupně, než je deg(f(x)) – definice D8.5.

Třídy modulo f(x)

S pomocí operace dělení polynomů budeme nyní definovat kongruenci dvou polynomů z množiny F[x]: Mějme dán nějaký f(x) ∈ F[x]. O polynomech a(x), b(x) ∈ F[x] říkáme, že jsou kongruentní modulo f(x) právě tehdy, když existuje q(x) ∈ F[x] tak, že a(x)-b(x) = q(x)*f(x). Tento vztah zapisujeme jako a(x) ≡ b(x) (mod f(x)) – definice D8.6.

Kongruence polynomů se tak definuje obdobným způsobem jako v případě celých čísel modulo n. Není složité ukázat, že kongruence dle D8.6 definuje na F[x] relaci ekvivalence. Volně řečeno ji tedy můžeme chápat jako “běžnou" relaci “rovná se". Přesnější však budeme, pokud si zavedeme pojem třída ekvivalence: Mějme dán polynom f(x) ∈ F[x]. Třída ekvivalence obsahující g(x) ∈ F[x] je definována jako množina [g(x)] = { h(x): h(x) ≡ g(x) (mod f(x)), h(x) ∈ F[x] } – definice D8.7.

Smysl zavedení tříd ekvivalence je pro nás v tom, že ačkoliv tyto množiny samy o sobě nejsou

konečné, množina všech tříd ekvivalence pro daný polynom $f(x) \in F[x]$ konečná je. Množinu všech tříd ekvivalence pro vybraný polynom $f(x) \in F[x]$ značíme $F[x]/f(x)$ – definice D8.8.

Není dále těžké ukázat, že každá třída ekvivalence obsahuje právě jeden polynom $g(x) \in F[x]$, pro který platí $\deg(g(x)) < \deg(f(x))$. Máme-li takový polynom, potom můžeme příslušnou třídu definovat jako $[g(x)] = \{ h(x) = g(x) + q(x)f(x): q(x) \in F[x] \}$. Tuto vlastnost je vhodné zdůraznit proto, že celou strukturu $F[x]/f(x)$ můžeme popsat pomocí všech polynomů stupně menšího než $\deg(f(x))$, čehož se s výhodou užívá při implementaci této struktury v HW a SW prostředcích. (Je to stejné jako v Z_p, ve kterém se zajímáme také pouze o čísla menší než p, ačkoliv bychom místo každého z nich mohli používat jakýkoliv jiný prvek z téže třídy.)

Věnujme se nyní zavedení operací sčítání a násobení na $F[x]/f(x)$. Tyto operace jsou zde definovány následujícím způsobem: $[a(x)] + [b(x)] = [a(x) + b(x)]$, $[a(x)] * [b(x)] = [a(x) * b(x)]$ - definice D8.9. Poznamenejme, že zatím se zde přísně držíme formální definice $F[x]/f(x)$, a proto zacházíme s jejími prvky jako se třídami. V běžné teorii se však mlčky toleruje zápis $g(x) \in F[x]/f(x)$, který chápeme ovšem jako $[g(x)] \in F[x]/f(x)$. (Viz ostatně opět zacházení se Z_p, kde se nad tím ani nepozastavujeme.)

Opět je snadné dokázat, že množina $F[x]/f(x)$ spolu s operacemi dle D8.9 tvoří okruh – tvrzení T8.7. Dále platí, že $F[x]/f(x)$ spolu s operacemi dle D8.9 je těleso právě tehdy, když je polynom $f(x)$ ireducibilní nad F – tvrzení T8.9. Zde můžeme spatřit jistou analogii mezi vlastnostmi užití ireducibilních polynomů a prvočísel.

Závěr

V tomto převážně algebraickém dílu jsme si ukázali základní struktury, které se v teorii ECC používají nejčastěji. Zobecnili jsme si přitom běžně známé pojmy, jako je operace sčítání a násobení, a ukázali jsme si způsobe konstrukce konečného okruhu/tělesa $F[x]/f(x)$. Příště se budeme věnovat způsobu využití této struktury pro konstrukci cyklických kódů.

Tomáš Rosa