POSUDEK OPONENTA NA BAKALÁŘSKOU PRÁCI
DAVIDA MAŘÁKA
KORELAČNÍ ÚTOKY

Práce se zabývá korelačními útoky na proudové šifry založené na lineárních rekurencích. Zmiňuje také vztah k dekódování samoopravných kódů. Tematicky student splnil zadání, které ovšem ve svázané kopii, kterou jsem měl k dispozici, chybí.

Přínosem práce mělo být zřejmě přehledné zpracování tématu korelačních útoků z různých zdrojů. Přínosem by bylo také objasnění podobnosti s dekódováním.

Text nenaplňuje očekávání kladená na bakalářskou práci, pokud jde o srozumitelnost, přehlednost a korektnost matematického textu. Značení je nedůsledné a nepřesné. Výklad je často nesrozumitelný nebo natolik neformální, že vyžaduje předběžnou dobrou znalost tématu. K přehlednosti nepřispívá ani tematické členění, příbuzná témata jsou pojednána na různých místech.

Upřesnění výtek.

- Souvislost mezi lineárními rekurentními posloupnostmi (LRP) a registry (LFSR) by neměla být jen předmětem poznámky. LRP je matematický popis LFSR a měl by být tedy důsledně využíván. Definice 3 vlastně není matematickou definicí, ale návodem k sestavení mechanického nástroje generujícího danou posloupnost.
- Nedostatečná identifikace LRP jako struktury se kterou pracujeme, se projevuje také v roztříštěnosti značení a fragmentárním zkoumání jejích vlastností. Viz např. Tvrzení 6, objevující se až na str. 16. Toto tvrzení je navíc formulováno nesprávně. Nejen že se namísto *linární kombinace* mluví o „součtu“, ale tvrzení je v uvedené podobě triviální („jedničku i nulu lze zapsat jako součet jedniček a nul“). Maticový zápis z kapitoly 4.1 je pak jen dalším vyjádřením téže vlastnosti. V čem tedy spočívá „vylepšení“? Vlastnost je upřesněna bezděky až v kapitole 4.2.
- Princip Siegenthalerova útoku si musí čtenář domýšlet. Celý výklad např. nevylučuje, že korelace je nulová. Funguje útok i v tomto případě? Implicitní zůstává i základní úvaha o tom, v čem je výhoda dešifrování registrů po jednom.
- Proč je korelace funkcí z Definice 18 dvojnásobkem korelace posloupností? Tato definice ostatně nedává smysl, dokud f a g nejsou náhodné veličiny.
- Podobně Věta 2 dává smysl, jen pokud jsou f a $l^{(u)}$ náhodné veličiny s rovnoměrným rozdělením. Pak je naopak její důkaz snadný a není nutné se odvolávat do literatury. (Toto je jedno z míst, kde si nejsem jist, zda student rozumí tomu, co píše).
- Kapitola 2.2 by měla obsahovat jednu z nejdůležitějších sdělení práce. Je však nejasná a nepřesná. Tvrzení 4 např. není formulováno srozumitelně. Co znamená „jednoznačné dekódování“ a v jakém smyslu je danou délkou zajištěno? Důkaz je zcela neformální a neměl by být vůbec nazýván důkazem. Lepší by bylo srozumitelně vysvětlit jeho myšlenku. Sdělení kapitoly se potom bez upozornění na souvislost opakuje v kapitole 4.2.
- Kapitola 3 dlouho nezmiňuje klíčovou otázku lineární nezávislosti hledaných rovnic. Zmínka se objeví až v kapitole 3.3 a jen v omezené souvislosti.
- Výklady o pravděpodobnostech v kapitole 3 je vinou značení a neuspořádanosti velmi obtížné sledovat. Jedná se přitom o poměrně jednoduché úvahy. Nejasná je již charakteristika nezávislosti množin U a Z.
- Formule (3.10) na str. 14 není „alternativní formule“ hodnoty $s(p,t)$, ale je řešením rekurence z Tvrzení 5. Je navíc uvedena chybně, jak lze snadno ověřit dosazením $t = 1$.
- Pro celou kapitolu 3 je typické, že se značí jednoduchou proměnnou hodnoty, které jsou funkcemi jiných hodnot. Jako ilustrativní příklad uveďme větu:

  > Jak vidíme z příkladu (3.2.1) opravdu platí, že pravděpodobnost p^* se zvýší pokud $z_n = u_n$, a sníží pokud $z_n \neq u_n$.

Příkladem (3.2.1) je míněn předchozí, nečíslovaný příklad? Jak je chování p^* z příkladu zřejmé? A může vůbec dávat toto tvrzení nějaký smysl, pokud dohledáme, že p^* značí $P(z_n = u_n \mid S)$? (Ve větě také chybějí tři čárky.)

Další poznámky.

- Pro definování pojmu se obvykle nepoužívá „právě tehdy“.
- Definované pojmy nejsou v práci graficky vyznačeny.
- Definice 10 a Definice 11 nejsou korektní. $f^{(m)}$ je množina funkcí? Co znamená zápis $\forall f^{(m)}$?
- Definice 15 obsahuje tvrzení.
- Informační entropie je definovaná pro prostor stavů, ale vzápětí se (pro binární případ) aplikuje na reálné číslo.
- Co znamená informace o efektivitě útoku na str. 9? Má „efektivní“ nějaký přesnější význam?
- Písmenem g se značí jak obnovovací funkce, tak charakteristický polynom.
- Není vyjasněno, s jakým tělesem pracujeme (někdy je to $\mathbb{F}_q$, jindy se automaticky předpokládá $q = 2$).
- Odkazy typu „Plyne z věty (2)“ jsou typograficky nesprávné.
- V práci nejsou nijak využity vlastnosti charakteristického polynomu (ireducibilita, primitivita). Proč jsou tedy definovány?
- Používání pojmů z pravděpodobnosti je nedůsledné. Je např. „nedeterministická funkce“ z Definice 6 náhodná veličina?
- Jsou formule (3.12) a (3.13) správně? Zdá se, že součet pravděpodobností není jedna.

Práce přinejmenším nesplňuje jedno z hlavních kritérií, totiž že má obsahovat alespoň jednu část, na níž student prokáže schopnost prezentovat rigorózním a korektním způsobem matematický text. Proto ji nedoporučuji k obhajobě.

Praha 18. srpna 2014 Štěpán Holub