Pokroky matematiky, fyziky a astronomie

Lucie Kárná; Štěpán Klapka

Bezpečnostní kódy v železničních zabezpečovacích zařízeních

Pokroky matematiky, fyziky a astronomie, Vol. 58 (2013), No. 2, 100–106

Persistent URL: http://dml.cz/dmlcz/143376

Bezpečnostní kódy v železničních zabezpečovacích zařízeních

Lucie Kárná, Štěpán Klapka, Praha

1. Úvod

Bezpečnost zabezpečovacích zařízení, jako jsou například železniční přejezdy, se skládá z několika aspektů. V tomto článku se zaměříme na *bezpečnost komunikace*, a to jak komunikace mezi jednotlivými prvky systému, tak mezi systémy vzájemně (resp. mezi zabezpečovacím zařízením a jeho okolím).

Pojem *bezpečnost* je poněkud kostrbatě definován jako „nepřítomnost nepřijatelných úrovní rizika“. Pro naše účely postačí, budeme-li bezpečnost chápat v jejím běžném smyslu. Bezpečnost systému má dvě složky: *funkční bezpečnost*, což je způsob reakce systému na různé kombinace vnějších vstupů a vnitřních stavů systému (tj. „co to dělá“) a *integritu bezpečnosti* – to je schopnost systému požadované funkce skutečně vykonávat. Integrita bezpečnosti se týká odolnosti systému jak vůči systematickým, tak vůči náhodným poruchám. Přitom pouze požadavky na integritu vůči náhodným poruchám jsou kvantifikovatelné.

V tomto článku se tedy budeme zabývat jedním z modelů vlivu náhodných poruch na integritu bezpečnosti, konkrétně na bezpečnost komunikace.

1.1. Bezpečná komunikace

Co vlastně znamená „bezpečnost komunikace“? Bezpečná komunikace musí zaručit, že zpráva pochází ze stanoveného zdroje (*autenticita* zpráv), že doručená informace je kompletní a nezměněná (*integrita*), a že jsou zprávy doručeny ve správném čase (*včasnost*) a ve správném pořadí (správné *řazení*). V některých případech je jako další bezpečnostní služba požadováno zajištění *důvěrnosti*, což znamená, že informace nesmějí být přístupné neoprávněným subjektům.[1]

Existuje řada technik, které se k zajištění uvedených *bezpečnostních služeb* používají: zpráva může být rozšířena o pořadové číslo (sekvenční značku), časový údaj nebo indikátor zdroje a adresáta, může být prováděna kontrola maximální délky časové prodlevy mezi dvěma zprávami, příjemce může odesílateli vysílat zpětnou zprávu

[1]Součástí zabezpečení informací obvykle bývá i dostupnost zprávy na vyžádání oprávněným uživatelem. V případě bezpečné komunikace na železnici se však přijímač při nedostupnosti potřebné informace musí zachovat tak, aby zajistil bezpečný stav zařízení. Proto zde dostupnost informace není podmínkou bezpečnosti a má význam jen pro spolehlivost systému.

Mgr. Lucie Kárná, Ph.D., Ústav aplikované matematiky, Dopravní fakulta ČVUT, Na Florenci 25, 110 00 Praha 1, e-mail: karna@fd.cvut.cz, doc. RNDr. Štěpán Klapka, Ph.D., Věda a výzkum, AŽD Praha s. r. o., Žirovnická 2, 106 17 Praha 10, e-mail: klapka.stepan@azd.cz

Obr. 1. Slavná nehoda na pařížském nádraží Montparnasse roku 1895. Strojvůdce začal brzdit příliš pozdě (zřejmě také nevěděl, že záchranná brzda je z úsporných důvodů odpojená) a vlak projel celou nádražní halu, prorazil její čelní stěnu a vypadl až na bulvár před nádraží. Jedinou obětí byla žena, která před nádražím prodávala noviny.

s různým obsahem, může být zaveden složitější postup identifikace účastníků komunikace, zpráva může být zajištěna bezpečnostním kódem nebo mohou být použity různé kryptografické techniky. Protože každý z těchto postupů poskytuje ochranu proti různým základním chybám, používá se zpravidla kombinace několika z nich.

2. Bezpečnostní kódy

Jako *bezpečnostní kód* se označuje kód umožňující odhalit určité typy chyb ve zprávě, je-li použit v bezpečnostně relevantní aplikaci za účelem zajištění její bezpečnosti. Bezpečnostní kód má mezi obrannými technikami zvláštní postavení, protože je jediným způsobem ochrany proti poškození zprávy. Proto je na příslušné protokolové vrstvě použití bezpečnostního kódu nezbytné. Mezinárodní standardy pro různé typy systémů vyžadují jeho povinné použití (pro železniční aplikace je to norma [1]).

2.1. Binární lineární kódy

V technické praxi jsou používány téměř výhradně kódy binární. Používáme běžné značení $\mathbf{Z}_2 = \{0, 1\}$ pro těleso zbytkových tříd modulo 2. Jako *(binární) slovo délky n* označíme libovolný vektor délky n, jehož složky jsou prvky ze $\mathbf{Z}_2$ – tedy jedničky a nuly.

Binární lineární (n, k)*-kód* $\mathbf{K}$ je pak libovolný podprostor prostoru $\mathbf{Z}_2^n$ s dimenzí k, to znamená množina vektorů nul a jedniček délky n, která je uzavřená vzhledem k operaci sčítání. (V tělese $\mathbf{Z}_2$ z toho již vyplývá i uzavřenost vzhledem k násobení skalárem.)

Vektory z prostoru $\mathbf{Z}_2^n$ se tradičně označují jako *slova*. Slova z kódu $\mathbf{K}$ se nazývají *kódová slova*, ostatní prvky prostoru $\mathbf{Z}_2^n$ jsou *nekódová slova*.

Dimenze k kódu odpovídá počtu *informačních bitů* kódu, tj. počtu bitů, které v každém slově reprezentují přenášenou informaci. Rozdíl $n - k$ je počet *kontrolních* neboli *redundantních bitů* kódu; to jsou bity, které kód k informaci přidává pro kontrolní účely.

Nejjednodušším příkladem binárního lineárního kódu je kontrola parity. Kód *sudá parita* se skládá ze všech slov dané délky n se sudým počtem jedniček. Tento kód má $n - 1$ informačních bitů a 1 kontrolní bit. Kontrola sudé parity je použitá jako bezpečnostní kód ve většině hardwarových i softwarových aplikací.

2.2. Detekce chyb

Při přenosu zakódované informace v prostoru (vysílání a přijímání zprávy) nebo v čase (uložení informace na paměťové médium a její čtení po delší době) může být zpráva různými vnějšími vlivy narušena. To se na úrovni jednotlivých bitů může projevit tak, že jeden či více bitů zprávy chybí nebo jsou naopak přidány nové bity (skluz synchronizace). Dále se budeme zabývat pouze typem poruchy, kdy je počet bitů zachován, ale dojde k záměně některého (některých) z nich za jiný (jiné).

Jako *detekční kód* označujeme takový kód, který tento typ poruch při přenosu objevuje. Princip detekce poruch je následující: Vysílající prvek vyšle kódové slovo u, přijímající prvek přijme nějaké slovo v ze $\mathbf{Z}_2^n$, které může a nemusí být kódové. Nyní mohou nastat dvě možnosti: buď je přijaté slovo v nekódové – pak je zřejmé, že toto slovo určitě nebylo vysláno a došlo k *chybě*, která je tímto objevena.

Druhá možnost je, že přijaté slovo v je kódové. To ovšem může odpovídat dvěma různým scénářům: buď je to právě to slovo, které bylo vysláno (to znamená, že přenos proběhl bez poruchy), nebo porucha při přenosu vytvořila jiné kódové slovo, než bylo původně vysláno. Protože přijímač nemá žádnou možnost rozeznat, který z těchto dvou scénářů nastal, je poslední eventualita nepříznivá – došlo k chybě, kterou kód neobjevil. Pravděpodobnost takové *neodhalené (nedetekované) chyby* je u kódů používaných v bezpečnostně kritických aplikacích (což je i řízení dopravy) velmi důležitý parametr.

Z uzavřenosti lineárního kódu jako množiny vzhledem ke sčítání (a tedy i k odčítání) plyne, že pro $u \neq v$ je chyba kódem odhalena právě tehdy, když rozdíl přijatého a vyslaného slova $v - u$ není kódovým slovem (viz například [2]). To je velká výhoda lineárních kódů, protože při výpočtu pravděpodobnosti nedetekované chyby nemusíme brát v úvahu všechny možné dvojice slov (vyslaného a přijatého). Stačí se omezit na situaci, kdy je vysláno slovo nulové, tj. se všemi složkami rovnými nule.

2.3. Váhový vektor kódu

Základním parametrem, ovlivňujícím schopnost kódu odhalovat chyby, je jeho *minimální vzdálenost*. Definujeme nejprve *Hammingovu váhu* slova jako počet jeho bitů, které jsou různé od nuly. *Minimální vzdálenost* lineárního kódu je nejmenší Hammingova váha nenulového kódového slova.

Kód s minimální vzdáleností d objevuje všechny chyby, při kterých je ve slově změněno nejvýše $d-1$ bitů ve slově. Neobjeví ale všechny chyby s d nebo více chybnými bity ve slově (viz [2]). Některé z nich však objevit může – které to budou, to závisí na detailnější struktuře kódu. Ta může být u lineárního kódu plně popsána *váhovým vektorem kódu* $A = (A_1, A_2, \ldots, A_n)$, kde A_i označuje počet kódových slov s Hammingovou váhou i.

2.4. Pravděpodobnost nedetekované chyby

Nyní odvodíme vzorec pro pravděpodobnost nedetekované chyby lineárního kódu. Ta je stejná jako pravděpodobnost, že při vyslání nulového slova bude přijato jiné (nenulové) kódové slovo (viz [2]). Předpokládejme tedy, že bylo vysláno nulové slovo, a že bylo přijato slovo, ve kterém je právě i jedniček, tedy chybných bitů. Pravděpodobnost, že toto přijaté slovo je kódové, vypočteme jako podíl počtu A_i kódových slov s Hammingovou váhou i a počtu všech binárních slov délky n, obsahujících i jedniček, což je $\binom{n}{i}$. Označíme-li P_i pravděpodobnost, že v přijatém slově je chybně právě i bitů, je potom pravděpodobnost P_{ud} nedetekované chyby kódu rovna

$$P_{ud} = \sum_{i=1}^{n} P_i \frac{A_i}{\binom{n}{i}}. \tag{1}$$

Pravděpodobnost P_i právě i chybných bitů ve slově ovšem závisí na podmínkách přenosu informace a nikoliv na vlastnostech samotného kódu. Mechanismus přenosu informace se popisuje pomocí *přenosového kanálu* a jeho vlastnosti mohou být rozmanité. Je poměrně obtížné nalézt takový model, který by alespoň zhruba odpovídal reálným podmínkám a přitom umožňoval výpočty vedoucí k nějakým závěrům.

3. Binární symetrický kanál

Nejčastěji využívaným modelem přenosového kanálu je *binární symetrický kanál bez paměti (BSC)*. Tento kanál má jak na vstupu, tak na výstupu prvky $\mathbf{Z}_2$ (jedničky a nuly) a je bez paměti, což znamená, že pravděpodobnost chybného přenosu bitu nezávisí na předchozích přenášených bitech. Třetí charakterizující vlastností BSC je jeho symetrie: pravděpodobnost p_e chybného přenosu bitu (*pravděpodobnost bitové chyby*) je stejná pro oba směry, to znamená jak pro změnu nuly na jedničku, tak pro změnu jedničky na nulu.

V binárním symetrickém kanále je zřejmě pravděpodobnost P_i, že ve slově délky n vznikne chyba právě v i bitech, rovna

$$P_i = \binom{n}{i} p_e^i (1-p_e)^{n-i}. \tag{2}$$

Jaká je tedy pravděpodobnost nedetekované chyby binárního lineárního kódu v binárním symetrickém kanálu? Dosadíme-li do vzorce (1) za pravděpodobnost P_i vztah (2), dostaneme

$$P_{ud}(p_e) = \sum_{i=1}^{n} {p_e}^i (1-p_e)^{n-i} A_i. \tag{3}$$

Zbývá ještě určit počty A_i kódových slov s Hammingovou váhou i. Binární lineární (n,k)-kód je lineární podprostor $\mathbf{Z}_2^n$ s dimenzí k, má tedy právě 2^k prvků. To znamená, že platí

$$A_0 + A_1 + \cdots + A_n = 2^k.$$

Každý lineární kód obsahuje nulové slovo, proto je $A_0 = 1$. Je-li minimální vzdálenost kódu rovna d, platí $A_1 = A_2 = \cdots = A_{d-1} = 0$.

Ostatní hodnoty A_i jsou známy pro několik málo typů speciálně zkonstruovaných kódů. U ostatních kódů je k jejich zjištění zapotřebí velmi pracný výpočet, vyžadující vygenerování 2^{n-k} kódových slov. (Blíže viz například [2].) Algoritmus výpočtu je sice jednoduchý a dá se snadno paralelizovat, ale při běžně používaných hodnotách $n-k$ (32, 48, 64 až 96) je i tak jeho časová náročnost enormní.

3.1. ‚Dobrý' a ‚správný' kód

Obtížnost výpočtu váhového vektoru kódu vede ke snaze najít nějaký lépe zvládnutelný postup určení pravděpodobnosti nedetekované chyby kódu. Především není nutné znát celý průběh funkce $P_{ud}(p_e)$; pro další bezpečnostní úvahy je postačující hodnota jejího maxima.

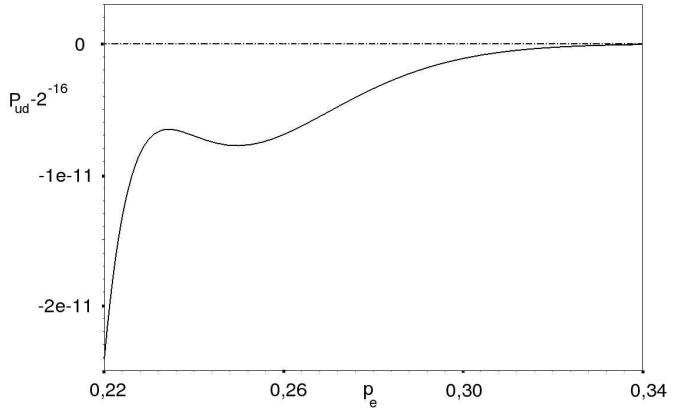
Dále nemusíme brát v úvahu všechny možné hodnoty pravděpodobnosti bitové chyby p_e. Kanál, pro který by hodnota p_e byla rovna jedné, by vysílanou informaci přesně invertoval a byl by tedy vlastně stejně „bezchybný" jako kanál s nulovou pravděpodobností bitové chyby $p_e = 0$. Obecně mají kanály s pravěpodobností bitové chyby větší než $1/2$ tendenci zprávu spíše invertovat než přenášet nezměněnou. Proto stačí, když kód poměrně jednoduchým způsobem ochráníme proti nedetekované inverzi a budeme uvažovat hodnoty p_e v intervalu $[0, 1/2]$.

Dosadíme-li krajní hodnotu $p_e = 1/2$ do vzorce (3) pro pravděpodobnost nedetekované chyby v BSC, dostaneme

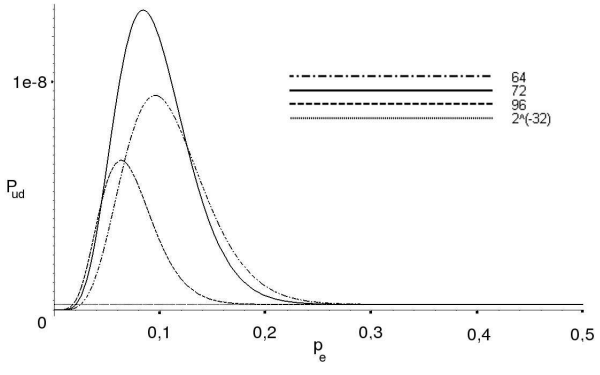
$$P_{ud}(1/2) = \frac{2^k - 1}{2^n} < 2^{k-n}.$$

Tato hodnota platí pro všechny binární lineární (n,k)-kódy, což je velmi praktické. Navíc na nějakém okolí bodu $1/2$ platí odhad $P_{ud}(p_e) < 2^{k-n}$. Pokud by tento odhad platil na celém intervalu $[0, 1/2]$, nemuseli bychom již kód podrobovat žádnému dalšímu zkoumání. To vede k následující definici[2] pojmů ‚dobrý kód' a ‚správný kód':

- Binární lineární (n,k)-kód je *‚dobrý' (good)*, jestliže pro všechna $p_e \in [0, 1/2]$ platí nerovnost $P_{ud}(p_e) < 2^{k-n}$.
- Binární lineární (n,k)-kód je *‚správný' (proper)*, jestliže je funkce $P_{ud}(p_e)$ monotónně rostoucí na intervalu $[0, 1/2]$.



Obr. 2. Příklad průběhu funkce $P_{ud}(p_e)$ na části intervalu $[0, 1/2]$ pro kód, který je ‚dobrý' (hodnota $P_{ud}(p_e)$ je na celém intervalu menší než 2^{k-n}), ale není ‚správný' (funkce $P_{ud}(p_e)$ není na celém intervalu rostoucí). Takových kódů jsme nalezli jen velmi málo. Aby měl obrázek dostatečné rozlišení, nejsou na svislé ose vyneseny přímo hodnoty funkce $P_{ud}(p_e)$, ale rozdíl $P_{ud}(p_e) - 2^{-16}$; proto jsou hodnoty záporné.



Obr. 3. Příklad kódů, které nejsou ‚správné'. Jedná se o tři různé kódy vzniklé zkrácením téhož cyklického kódu. Délka jednotlivých kódů je $n = 64$ až $n = 96$, počet kontrolních bitů je u všech kódů roven $n - k = 32$. Vodorovná čára u dolního okraje grafu je hodnota 2^{-32}; maximální hodnota pravděpodobnosti nedetekované chyby u nejhoršího z uvedených kódů je více než 50krát vyšší.

Zřejmě je ‚správný' kód vždycky také ‚dobrý' a zdálo by se zbytečné tento pojem vůbec zavádět. Jeho význam je v tom, že ‚správný' kód má na kanálech s menší chybovostí menší pravděpodobnost selhání než na horších (chybovějších) kanálech, což je chování, které bychom od rozumného detekčního kódu očekávali.

[2]Tyto termíny nejsou zrovna šťastně zvoleny. Vznikly překladem anglických termínů *good* a *proper*, které jsou také poněkud nepraktické. Držíme se terminologie používané v existující literatuře, především v české mutaci evropské normy [1].

Druhý důvod pro zavedení pojmu ‚správný' kód je, že se monotonie funkce $P_{ud}(p_e)$ dá pro některé třídy kódů obecně dokázat, například pro perfektní kódy, MDS kódy nebo některé BCH kódy (viz např. [3], [4]). Takové kódy jsou tedy také ‚dobré' a pravděpodobnost nedetekované chyby v BSC je shora omezená známou hodnotou 2^{k-n}, která může být použita v následných výpočtech bezpečnosti celého systému.

Odsud vzniklo všeobecně rozšířené přesvědčení, které by se dalo neformálně vyjádřit slovy: „Všechny slušné kódy jsou ‚správné'." Jelikož se již od 70. let minulého století objevovaly články upozorňující na běžně používané kódy, které ‚správné' nejsou (viz např. [4]), přibyl názor „I když nejsou ‚správné', tak se od ‚správných' kódů moc neliší a překračují hodnotu 2^{k-n} jen o málo."

Toto přesvědčení se však ukázalo být zcela mylné. Ukázalo se, že mnoho (ne-li většina) používaných kódů ‚správných' není a hodnotu pravděpodobnosti nedetekované chyby 2^{k-n} překračují často velmi výrazně, a to obvykle pro poměrně nízké hodnoty chybovosti kanálu. Nalezli jsme například kód, u kterého by se vzhledem ke způsobu jeho konstrukce daly očekávat poměrně příznivé detekční vlastnosti, ale jeho maximum pravděpodobnosti nedetekované chyby bylo více než tisíckrát vyšší, než hodnota 2^{k-n}.

4. Závěr

Pokud je nám známo, nebyla dosud zaznamenána žádná nehoda, způsobená selháním bezpečnostního kódu. Příčiny většiny železničních nehod jsou mnohem prozaičtější. Zatimco v minulosti byly velmi časté nepředvídatelné technické závady, dnes převažují závady způsobené zanedbáním předepsané údržby a nehody způsobené chybou obsluhy. V obou případech jde tedy o selhání lidského faktoru.

Současné trendy jsou charakterizovány přesunem k unifikovaným interoperabilním komunikačním rozhraním (na evropské i světové úrovni), která jsou většinou určena pro použití v otevřených přenosových systémech (evropské systémy ERTMS/ETCS, GSM-R, protokol EURORADIO). Bezpečnostní kódy v těchto systémech používají kryptografické techniky. Ty není možné hodnotit výše uvedeným postupem, nebo je toto hodnocení mimo možnosti současných výpočetních prostředků. Proto je předmětem současných diskusí, zda by platné normy měly nadále umožňovat propojení ochran pro zajištění integrity a autenticity. Toto spojení totiž generuje kontroverzní požadavky.

Dalším otevřeným problémem je zajištění nezávislosti bezpečnostního a přenosového kódu. Dosud totiž nepanuje shoda ani v tom, jak tuto nezávislost vlastně definovat.

Literatura

[1] *EN 50159 Drážní zařízení: Sdělovací a zabezpečovací systémy a systémy zpracování dat — Komunikace v přenosových zabezpečovacích systémech*, UNMZ, 2011.

[2] Huffman, W. C., Pless, V.: *Fundamentals of error-correcting codes.* Cambridge University Press, Cambridge, 2003.

[3] Kasami, T., Lin, S.: *On the probability of undetected error for the maximum distance separable codes.* IEEE Trans. Commun. *9* (1984), 998–1006.

[4] Leung-Yan-Cheong, S. K., Hellman, M. E.: *Concerning a bound on undetected error probability.* IEEE Trans. Inform. Theory *3* (1976), 235–237.