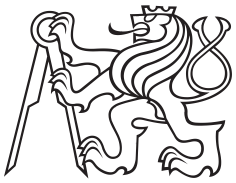
Diplomová práce



České
vysoké
učení technické
v Praze

F3

Fakulta elektrotechnická
Katedra telekomunikační techniky

Přehledový přijímač / monitor rádiových sítí IoT

Ondřej Šulc

Školitel: Ing. Pavel Troller, CSc.
Obor: Komunikační systémy a sítě
Leden 2019

Poděkování

Děkujeme ...

Prohlášení

Fakt sám ...

Abstrakt

Rozvíjíme ...

Klíčová slova: IoT, SDR-RTL, LoRa, Sigfox, Přehledový přijímač

Školitel: Ing. Pavel Troller, CSc.
Pestitelský ústav,
Zárivá 232,
12000 Praha 2

Abstract

We develop ...

Keywords: IoT, SDR-RTL, LoRa, Sigfox, Scanner

Title translation: Scanner/Monitor of IoT radio networks

Obsah

Obrázky Tabulky

Kapitola 1
Úvod

Foo bar

Kapitola 2
LoRa

2.1 Fyzická vrstva (LoRa PHY)

2.1.1 Modulace

Modulační schéma LoRa je založeno na Chirp Spread Spread Spectrum (Cvrlikající rozprostřené spektrum) modulaci (Goursaud and Gorce, 2015) a definuje jeden "cvrk" jako jeden symbol (Semtech, 2015a). Standardní nemodulovaný lineární cvrk se nazývá "základní cvrk" a může být matematicky popsán jako funkce času t takto (Mann and Haykin, 1991):

$$x(t) = e^{i(\varphi_0 + 2\pi(\frac{k}{2}t^2 + f_0 t))} \tag{2.1}$$

Kde φ_0 je počáteční fáze, k je rychlost změny frekvence a f_0 je počáteční frekvence. Pokud je šířka pásma kanálu BW, tak parametry f_0 a k jsou nastaveny tak, že se frekvence zvětšuje od $f_0 - \frac{BW}{2}$ po $f_0 + \frac{BW}{2}$ během periody T cvrku. Tím pádem je $f_0 = \frac{BW}{2}$ and $k = \frac{BW}{T}$. Doba trvání jednoho cvrku závisí na šířce pásma signálu a na parametru nazývaném činitel rozprostření (Spreading Factor - SF) dle vztahu $T = \frac{2^{SF}}{BW}$ (Seller and Sornin, 2014). Vzhledem k tomu, že $x(t + nT) = x(t)$ kde $n \in \mathbb{N}$, celočíselná hodnota $i \in \{0, 1\}^{SF}$ může být namodulována na základní cvrk pomocí časového posunu $\hat{t} = Gray^{-1}(i)\frac{T}{2^{SF}}$ aplikovaného na signál ve vztahu (**??**), kde $Gray^1$ je dekódování Grayova kódu (Gray, 1953). Touto cestou je symbol v podstatě kvantovaný na 2^{SF} časových intervalů rozdělujích šířku pásma, nazýváme je "chipy" a právě ony určují i. Při příjmu modulovaného cvrku s neznámým časovým posuvem $x(t + \hat{t})$, může být hodnota cvrku zrekonstruována navzorkováním signálu vzorkovací frekvencí chipů a výpočtem:

$$i = Gray(arg\max(|FFT(x(t + \hat{t}) \odot \overline{x(t)})|)) \tag{2.2}$$

Kde $\overline{x(t)}$ značí komplexně sdružený základní cvrk, $\odot$ značí multiplikaci po prvcích, $|FFT(x)|$ zančí velikost Rychlé Fourierovi transformace x, a $Gray$ je Grayovo kódování.

2.1.2 Prokládání

Jako v každé jiné modulaci, musíme i zde počítat s chybami způsobenými šumem, interferencí, a časovými nebo frekvenčními posuny. Tyto chyby mohou způsobit, že hodnota čipu nebude dobře odečtena z modulovaného symbolu. Například poryv šumu může posunout vrchol v FFT spektru na jinou hodnotu chipu a tak jej znehodnotit.
Aby bylo možné minimalizovat dopad poryvů šumu na chybu jen jednoho bitu v symbolu je použito prokládání. Několik chipů je dohromady vepsáno do mřížky $\{0,1\}^{SFx(4+CR)}$, kde CR (Coding Rate) značí počet paritních bitů a nabývá hodnot 1 až 4. Pokud tedy bude použit $SF = 7$ a $CR = 4$ dostaneme matici $\{0,1\}^{7x8}$, příklad je na obrázku **??**. K získání kódové slova je pak potřeba číst bity po diagonále matice. Na rozdíl od patentu LoRy (Seller and Sornin, 2014), kde se uvádí, že směr diagonálního čtení bitů z mřížky je směrem dolů, v praxi lze pozorovat opačný směr. Tímto způsobem tak první chip obsahuje všechny nejméně významné bity (LSB - Least significant) všech kódových slov, druhý čip všechny druhé bity všech slov a tak dále. Díky tomu v případě ztráty celého čipu dojde k chybě jen v jednom bitu na kódové slovo.
Dalším způsobem jak zvýšit odolnost proti rušení vysílání je použití módu redukované rychlosti (reduced rate mode). V případě použití tohoto módu jsou první dvě řady prokládací matice zahozeny a její rozměr se tak změní na $\{0,1\}^{SF-2x(4+CR)}$ což způsobí, že z ní nasledně vyčteme o dvě kódová slova méně. Zahozené řádky obsahují nejméně významné bity chipů, které jsou nachylnější k chybám protože odpovídají užším frekvenčním intervalům v FFT spektru. Z toho vyplývá, že mód redukované rychlosti obětuje rychlost přenosu dat ve prospěch odolnosti proti šumu. Hlavička fyzické vrstvy LoRa je v tomto módu vysílána vždy, kdežtkoo užitečná data je v případě použití SF 11 nebo 12.

2.1.3 Kódování

Po přečtení kódových slov z prokládací matice mají tato délku $4 + CR$. Kvůli zamezení vzniku stejnosměrné složky byla slova v části rámce s užitečnými daty XOR-ována 9-bitovým lineární posuvným registrem se zpětnou vazbou (LSFR Linear feedback shift register) (whitening). A proto musí po synchronizaci projít stejným procesem znovu. Přesný algoritmus není v patentu určen a jeho výběr je tedy na každém výrobci zvlášť.
Na několika testovacích zařízeních **??** reverzním inženýrstvým zjistilo použité upraveného $4/(4 + CR)$ Hammingova kódu. Ve výsledku tak z každého kódového slova po dekódování získáme 4 bity dat. Ta jsou pak naparsována du struktury rámce lora.

2.1.4 Struktura rámce

Na fyzické vrstě LoRa definuje rámec jako strukturu složenou z následujících polí. Pole jsou uvedena ve stejném pořadí jako v rámci. (Semtech, 2015b, p. 27–29)

Preambule Sekvence základních cvrků, která slouží k časové a frekvenční synchronizaci. Počet cvrků není pevně dán.

Symboly synchronizace rámce Dva modulované cvrky co mouhou být použity pro identifikaci sítě. Hardwarový přijímač zahodí rámcec, které obsahují synchronizační symboly co neodpovídají jeho nastavení.

Symboly synchronizace frekvence Dva sdružené cvrky následované sdruženým cvrkem s periodoou $\frac{T}{4}$ určené pro přesnou frekvenční synchronizaci.

Hlavička (nepoviná) Hlavička obsahuje délku užitečných dat, použitou přenosovou rychlost, indikuje použití Cyklického redundantního součtu (CRC - Cyclic redunduncy check) a jendobajtovou kontrolní sumu hlavičky. Pro modulaci hlavičky je vždy použito $CR = 4$ a mód redukované rychlosti. Pokud hlavička vysílána není (implicitní mód) musí mít jak přijímač tak vysílač předem schodně nastavený CR a také zdali je použito CRC.

Užitečná data Pole o proměnné délce obsahující data vrstvy přístupu k médiu (MAC - Media access control) a případné dvoubajtové CRC těchto dat.

2.1.5 Struktura hlavičky

Délka hlavičky není ve specifikaci nikdy přímo určena. Lze jí však vydedukovat z toho, že hlavička je vždy vysílána v módu redukované rychlosti, má $CR = 4$ a SF minimálně 7. Z toho vyplívá že hlavička se musí vejít do mřížky $\{0, 1\}^{7-2x8)}$ a to odpovídá 5 kódovým slovům. Každé slovo má 8 bitů a dohromady je to bitů 40. Jakékoliv zbývající bity jsou použity pro užitečná data.
Po dekódování díky redundantním bitům dostáváme $40\frac{4}{8} = 20$ bitů nebo 2,5 bajtu. V **??** experimentálně vyzkoušeli pořadí hlavičky. První bajt udává délku datového obsahu, následuje půlslabika udávající CR a přítomnost MAC CRC a poslední bajt obsahuje kontrolní součet hlavičky, z něj je však používá jen 5 LSB bitů.

2.2 Softwarová demodulace

?? dokázali implementovat kompletní PHY vrstvu LoRa ve frameworku GNU Radio. Jejich zdrojové kódy jsou open source a dostupné na Githubu. Funkčnost příjmu signálu LoRa mého scanneru vychází z jejich práce. V této kapitole je pospsán princip fungování.

2.2.1 Detekce a synchronizace

Aby mohl být signál demodulován musí být nejdříve detekován. K tomu slouží preamule která má dva opakující se cvrky čehož dokáže využít použitý

Schmidl-Cox algoritmus. Ten definuje dvě veličiny $P(d)$ a $R(d)$, ty jsou definované takto (Schmidl a Cox, 1997) **??**:

$$P(d) = \sum_{m=0}^{L-1} (x^*_{t+m} x_{t+m+L}) \tag{2.3}$$

$$R(d) = \sum_{m=0}^{L-1} |x_{t+m+L}|^2 \tag{2.4}$$

kde L je délka symbolu, t je index vzorku komplexnímo signálu x a x^* je jeho komplexně sdružený signál. Veličiny $P(d)$ a $R(d)$ jsou použity k výpočtu časové metriky $M(d)$:

$$M(d) = \frac{|P(d)|^2}{R(d)^2} \tag{2.5}$$

Časová metrika $M(d)$ v podstatě počítá normalizovanou autokorelaci délky L přes dva symboly, maximum bude mít ve cvhíli kdy v signálu budou za sebou dva totožné symboly. Díky tomu, že oba symboly jsou chybami způsobenými přenosem (interference, frekvenční odchylka nosné (CFO - Carrier frequency ofset), odchylka vzorkovací frekvence) ovlivněny stejně, tak tyto chyby téměř neovlivní výsledek korelace. Aby bylo možné efektivně počítat rovnice **??** a **??** v programu byla použita knihovna VOLK (Vector Optimized Library of Kernels), která implementuje SIMD (Single Instruction, Multiple Data) instrukce. Na obrázku **??** je vidět příklad výsledku použití časové metriky $M(d)$ na komplexním signálu LoRa. Kolem vzorku 2500 funkce dosahuje horní plošilny a poukazuje tak na existenci preambule.
I přesto že tento algorimus detekuje preambuli velmi dobře, není možné přesně určit počátek symbolu jen z horní plošiny časové metriky. Tým kolem Wanga **??** proto navrhl vylepšení kdy je od časové metriky $M(d)$ odečtena její opožděná verze $M_2(d)$, tím se z plošiny stává vrchol **??** a lze tak za začátek symbolu považovat vzorky odpovídající maximu této metriky. Nicméně ani to není jak je patrné z **??** dostatečně přesné pro signály LoRa.
Aby **??** tenhle problém vyřešili museli vymyslet nové řešení. To se zakládá použití Schmidl-Coxovi metriky pro přibližné určení okna ve kterém se nachází druhý symbol preambule a následném zpřesnění pomocí ideálního lokálně vygenerovaného cvrku. Jeho okamžitá frekvence $\omega_l(t)$ a normovaná okamžitá frekvence signálu Lora $\omega(t)$ jsou vzájemně korelovány (přes posuvné okno?) a index vzorku jež odpovídá maximální hodnotě této funkce je považován za počátek symbolu. Použití omažité frekvence místo komplexních hodnot je odůvodněno chybami CFO, které by bez korekce mohly ovlivnit přesnost synchronizace. Použitím okamžité frekvence jsou podobně jako u Schmidl-Coxova algoritmu tyto chyby zanedbatelné.

$$symbolstart = \underset{i \in \{0,1,\dots L\}}{\operatorname{argmax}} (\omega_l \star \omega)(i) \tag{2.6}$$

Výsledek je na **??**. Poslední součástí tohoto řešení je určení prahové hodnoty maxima korelačního koeficientu okamžitých frekvencí lokálnlně gerovaného cvrku a přijátého. Pokud je tato hodnota menší než prahová je daný rámec zahozen, protože se buďto jedná o falešně pozitivní detekci rámce nebo o nepovedenou synchronizaci.

2.2.2 Demodulace

Po úspěšné synchronizaci následuje fáze demodulace popsaná v předchozí kapitole. Oproti teorii má však v praxi FFT demodulace nevýhodu v tom, že je citlivá na odchylku frekvence, která způsobuje posun hodnot FFT a tím i odečítaných hodnot chipů. Tím pádem je potřeba přesná synchronizace frekvence, kterou je navíc potřeba aplikovat na každý kanál LoRa zvlášť. Separace kanálů a následná synchronizace každého z nich je však v softawaru příliš náročná operace a tak **??** přišli s novou metodou demodulace, která je nezávislá na frekvenci a umožňuje demodulaci na všech kanálech současně v reálném čase. V porovnání s FFT metodou je však méně robustní.
Nejdříve je potřeba spočítat okamžitou úhlovou fekvenci $\omega[t] = \frac{d\varphi[t]}{dt}$. Poté je potřeba $\omega[t]$ vyhladit a decimovat konstantním decimačním faktorem $\frac{s_f T}{2^{SF}}$ kde s_f je vzorkovací frekvence. Díky tomu je pak počet vzorků v $\omega[t]$ schodný s 2^{SF} následně je vypočítán digitální gradient f:

$$D_t[\omega[t]] = \omega[t+1] - \omega[t] \tag{2.7}$$

Tuto operaci si lze představit jako filtr horní propusť okamžité frekvence nebo jako druhou derivaci fáze. Protože frekvence základního cvrku se lineárně zvyšuje s k - $\omega(t) = kt + f_0$ je její derivace $\omega'(t)$ rovna k. Pro modulované cvrky se však v D_t objeví ostré šličky v místech přechodu mezi vysokou a nízkou frekvencí. Přítomnost takových špiček indikuje časový posun $\hat{t}$, nepřítomnost naopak idikuje časový posun 0 - základní cvrk.

Dalším problémem při demodulaci je zpožďování/předbíhání hodin v jednotlivých zařízeních. Kristalové oscilátory v LoRa vysílači a SDR se budou zákonitě navzájem předbíhat nebo zpožďovat, rozdíl jejich frekvencí je předem neznámý, ale v průbhu času se musí projevit. To může způsobovat problémy zejména v případě delšího datového obsahu v kombinaci s vyšším SF. V patentu LoRa jsou pro účely korekce tohoto jevu použity pilotní symboly, které pomohou sledovat časování. Ve skutečnosti se však zdá, že k jejich použití nebylo přistoupeno. Je tedy nutné využít techniku slepého odhadu, která využívá převzorkování přijatého signálu N-krát. Aby tato technika byla funkční je potřeba aby hodnota N odpovídala následujícímu vztahu $|\Delta t| < \frac{N}{2}$, kde Δt je chyba časování na symbol. Prvním krokem je synchronizace popsaná v **??**. Pokud je chyba časování na symbol $|\Delta t| < \frac{N}{2}$
, lze $|\Delta t|$ určit následujícím způsobem:

1. Symbol je demodulován běžným způsobem jak je popsáno v 2.2.2 a je tak získána hodnoto chipu i a časového posunu $\hat{t}$.

2. Na přijímači je lokálně generovaný základni cvrk modulován na i což způsobí časový posun $\hat{t}_1$ lokálního signálu.

3. Protože lokálně generovaný cvrk není ovlivněn rozdílem oscilátorů vysílače a přijímače můžeme vzájemné zpoždění oscilátorů definvat takto $\Delta t = \hat{t}_1 - \hat{t}$. Nyní stačí na příjimači opravit $\hat{t}$ připočtením vzájemného zpoždění k přijatému signálu.

Již zmíněná podmínka $|\Delta t| < \frac{N}{2}$ je daná tím, že při jejím nesplnění dekodér špatně určí hodnotu chipu i a chyba se tak bude šířit i do dalších symbolů. Hodnota N tak není určena přímo ale jako interpolace z $\hat{t}$ do $\hat{t} + \Delta t$. Vyšší hodnoty N by dále vylepšovali přesnost korekcí zpoždění ale také by zvyšovali náročnost výpočtu.

2.2.3 Dekódování

Ve fázi dekódování jsou hodnoty chipů zpětně proloženy a tím jsou získána kódová slova s 4 + CR bity. Prvních 8 kódových slov lze přímo dekódovat jako hlavičku fyzické vrstvy. Datový obsah však musí být nejprve XORován bělící posloupností. I přesto že dle výrobce LoRa Semtech má jít o sekvenci generovanou 9-bitovým LFSR ve skutečnosti je dle výzkumu **??** použita posloupnost mírně odlišná, která však není veřejně zdokumentovaná.
Pokud však známe původní kódové slovo i přijaté vybělené lze snadno zjistit sekvenci použitou pro XORování následovně $w^{(j)} = c_w^{(j)} \oplus c^{(j)}$. Pro zjednodušení lze vyslat všechna kódová slova nulová a tím dostaneme rovnici $w^{(j)} = c_w^{(j)} \oplus 0$ a na výstupu tak máme přío samotnou bělící sekvenci, kterou můžeme uložit a poté náčítat z tabulky.
Po Odbělování přichází poslední krok a to Hammingovo dekódování kódových slov. U LoRa jsou datové bity umístěny na jiných pozicích než je běžné a to na indexech 0,1,2 a 3 bajtu místo indexů 1,2,3 a 5. Graficky je toto mapování zobrazeno na **??**. Po extrakci datových bitů jsou pak pomocí paritních bitů detekovány a případně opraveny chyby a tím získána původní data.

Kapitola 3

Závěr

Lorep ipsum [1]

Literatura

[1] J. Doe. *Book on foobar.* Publisher X, 2300.