Formát BNL souboru

jindroush@seznam.cz, leden 2022

Co obsahuje BNL soubor

Veškeré informace níže jsou jen teorie ověřené na autorovi známých souborech, bez zkoumání zpracování firmwarem tužky.

- Číslo knihy – podle ťuknutí na tlačítko zapínání pak tužka vybere správný soubor
- Módy knihy – mezi módem 0 a 1 se přepíná dvojitým ťuknutím na stejný OID kód, na další módy pak výběrem speciálních ikon. Mód určuje, který zvuk se pro OID kód bude přehrávat.
- Vazbu mezi OID kódem a zvukovými soubory. Pro každý OID kód je tolik vazeb, kolik je módů knihy
- Kvízy – opačná vazba mezi zvukem otázky a OID kódem odpovědi
- Systémové globální tabulky zvuků pro kvízy, zvuky na systémových tlačítcích apod.
- Zvukové soubory MP3

Pořadí bloků v souboru

Pořadí jednotlivých bloků odpovídá tomu, jak jsou soubory uspořádány v distribuovaných BNL souborech.

- Hlavička obsahující ukazatele na další struktury a nějaké konstanty
- Tabulka offsetů (převodů OID kódů na zvuky)
- Tabulka kvízů, která svazuje intro kvízu s otázkami a správnými odpověďmi
- Tabulky OIDů
- Tabulky zvuků
- Tabulka offsetů na MP3 soubory
- Zašifrované MP3 soubory

Číselné formáty

Číselné formáty použité v tomto souboru jsou 8bitové (dále BYTE), 16bitové little-endian (dále WORD) a 32bitové little-endian (dále DWORD). Všechny ukazatele jsou absolutní DWORD od začátku souboru. Předpoklad je, že všechna čísla jsou unsigned (neznaménková).
OID je interní kód tužky, ať už převedený z raw kódu čtečkou z teček z papíru nebo jako virtuální kód použitý ve kvízu. Je to vždy WORD.
MediaId je index do tabulky MP3 souborů, od 0, WORD.

Hlavička

Ve všech zkoumaných souborech měla hlavička délku 0x200.

Offset	Typ	Popis
0x0000	DWORD	Header_key - Tímto se XORují všechny ukazatele v hlavičce, nejvyšší byte hodnoty má dále použití u derivace šifrovacího klíče MP3 souborů. Ačkoli pro funkčnost „čtení“ může být zcela nulový, nefungují kvízy. Podle pokusů se zdá, že stačí, aby byla nenulová hodnota na druhém bajtu zespoda, tj. 0x0000HH00. Teorie: je možné, že je potřeba k inicializaci random seedu pro kvíz?
0x0004	DWORD	? – délka hlavičky a/nebo ukazatel na první hodnotu tabulky OID kódů. Ve všech zkoumaných souborech 0x200. Při pokusech o změnu této hodnoty

		tužka nikdy nerozpoznala výsledný soubor (pokusy byly o plus mínus DWORD, na 0x300 a na 0x400).
0x0008	DWORD	Ukazatel na tabulku offsetů na MP3 soubory
0x000C	DWORD	Ukazatel na tabulku zvuků – přehrávána na první dotyk Start tlačítka
0x0010	DWORD	Ukazatel na tabulku zvuků – přehrávána na druhý dotyk Start tlačítka
0x0014	DWORD	Ukazatel na tabulku zvuků – neznámý účel
0x0018	WORD	V oficiálních souborech vždy 0. Z pokusů vyplynulo, že se jedná o nejnižší prvek v tabulce offsetů, tj. bázi tabulky offsetů.
0x001A	WORD	Poslední použitý OID kód. Délka tabulky offsetů je tím pádem (poslední-báze+1).
0x001C	WORD	Počet použitých media souborů – bohužel nesedí úplně přesně, většinou je o něco menší, než je počet skutečných a využitých MP3 souborů. Tužce však podle pokusů nevadí nastavení na „skutečný“ počet MP3 souborů.
0x001E	WORD	? – vždy 0. Pokus neprokázal, že by jakákoli hodnota zde měla vliv. Pravděpodobně padding předchozího WORDu.
0x0020	DWORD	Ukazatel na tabulku zvuků – neznámý účel
0x0024	DWORD	Ukazatel na tabulku zvuků – zvuk, který se ozve při přepnutí módu
0x0028	DWORD	? – vždy FFFFFFFF
0x002C	DWORD	Book_mode_read - Počet módů, které kniha podporuje (pravděpodobně jen WORD, horní WORD vždy 0 – padding?). Módy jsou přepínány tlačítky jako knížka, žárovka, bublina, vážou jeden OID k více zvukům.
0x0030	DWORD	? – 5x FFFFFFFF
0x0044	DWORD	Ukazatel na tabulku kvízů
0x0048	DWORD	Ukazatel na tabulku OIDů (přehrají se při správné odpovědi na kvíz)
0x004C	DWORD	Ukazatel na tabulku OIDů (přehrají se při správné odpovědi na kvíz)
0x0050	DWORD	Ukazatel na tabulku OIDů (přehrají se při špatné odpovědi na kvíz)
0x0054	DWORD	Ukazatel na tabulku OIDů (přehrají se při špatné odpovědi na kvíz)
0x0058	DWORD	Ukazatel na tabulku OIDů – neznámé použití
0x005C	WORD	Unikátní číslo knihy – raw kód pro tento OID je vytištěn na tlačítku Start.
0x005E	WORD	0, padding předchozího WORDu. Pokusem zjištěno, že nezávisí na jeho hodnotě.
0x0060	DWORD	15x ukazatel na tabulku zvuků s neznámým použitím
0x009C	DWORD	3x Ukazatel na tabulku OIDů s neznámým použitím
0x00A8	DWORD	Ukazatel na tabulku OIDů, nejčastěji s délkou 6. Obsahuje mluvené výsledky kvízu (od žádného dobře, až po všech 5 dobře)
0x00AC	DWORD	8x DWORD 0xFFFFFFFF
0x00CC	DWORD	29x DWORD ukazatele na tabulky médií s neznámým použitím
0x0140	DWORD	Pravděpodobně použito pro derivaci klíče, přesně neznámo. Hodnota nejnižšího bajtu tohoto DWORDu sečtená s nejvyšším bajtem header_key musí být 0xF5
0x0144	BYTE	16 bajtů tvořících klíč k dešifrování MP3 souborů. Klíč se získá tak, že se ke každému z těchto bajtů přičte nejvyšší bajt header_key
0x0154	DWORD	Až po offset 0x200 samé 0xFFFFFFFF

Tabulka offsetů

Pouze řídký seznam DWORDových ukazatelů - 0xFFFFFFFF je „prázdná“ hodnota. Každý ukazatel ukazuje na místo, kde je book_mode_read tabulek zvuků za sebou. Každá tabulka má 0 nebo 1 zvuk, nezahlédl jsem žádnou s větší délkou. Pokusem zjištěno, že více zvuků v tabulce prostě znamená postupné přehrávání zvuků. OID je offset do této tabulky, tím váže OID+mód ke zvuku. Od OIDu se

odečítá báze tabulky (0x18 v hlavičce).
Módy 0 a 1 se střídají, ať už mód 0 je aktivován automaticky po zapnutí nebo po stisknutí ikony „knížky“ – většina nebo všechny knihy mají vždy 0 a 1 mód shodný. K přepnutí mezi módy 0 a 1 dojde při opětovném načtení stejného OIDu.

Tabulka OIDů

Jednoduchá tabulka, která obsahuje WORDový počet položek N a za ním N WORDů OIDů.

Tabulka zvuků

Jednoduchá tabulka, která obsahuje WORDový počet položek N a za ním N WORDů zvuků.

Tabulka offsetů na MP3 soubory

Tabulka pro N MP3 souborů - pouze N+1 DWORDových offsetů na začátek MP3 souboru, poslední DWORD je délka celého souboru (a tím i konec posledního MP3 souboru). Začátky MP3 souborů jsou zarovnány na násobky 512, výplň jsou 0. Konec není zarovnán.

Tabulka kvízů

Nejkomplikovanější a nejméně probádaná část BNL souboru, v této jsou zatím největší nejasnosti:

Začíná seznamem DWORD ukazatelů na jednotlivé hlavičky kvízů. Není známa délka této tabulky, hned za ní následuje první hlavička kvízu, tím se dá procházení ukončit. Počet kvízů většinou odpovídá počtu dvoustran knihy, někdy násobeno více obtížnostmi kvízů.

Hlavička kvízu

0x0000	WORD	Neznámá hodnota, většinou 0. Hodnota 4 znamená speciální typ kvízu, viz níže. Jednou zahlédnuta hodnota 0x100, 2x hodnota 1. Hodnota 1 u „normálně“ konstruovaného kvízu pero pouze přečte úvod kvízu, ale nepokládá otázky. 0x100 se chová jako 0x00 (uvažuje se pouze dolní byte?)
0x0002	WORD	Qcnt – Počet možných otázek kvízu
0x0004	WORD	Počet položených otázek kvízu. Je nutno křížově kontrolovat, že pole na offsetu 0xA8 má vždy délku o jedna větší než je hodnota zde (Všechny špatně až po všechny dobře, N+1 možností). Teorie: podle všeho z toho plyne, že všechny kvízy musí mít nastaveno stejný počet otázek – jinak by nefungovalo správně vyhodnocení, některé knihy toto mají nastaveno různě (nemám k dispozici)
0x0006	WORD	Neznámá hodnota, většinou 0, občas 1, jednou 2 nebo 5. V případě, že je 0, stačí najít jeden z OIDů v poli otázek, pokud je 1 nebo 2, chce všechny OIDy, aby byla odpověď uznaná za správnou. Hodnota 3 neustále cykluje otázku.
0x0008	WORD	OID vedoucí na zvuk úvodu kvízu. Neplatná hodnota stejně způsobí spuštění správného kvízu bez úvodního zvuku. Teorie: znamená to tedy, že pořadí kvízů je dáno čistě pořadím v tabulce (a tím pádem je napevno nakódováno od 100) a OID je tam pouze jako vazba k úvodnímu zvuku?
0x000A	DWORD	Qcnt krát DWORD ukazatel na jednotlivé otázky

Jednotlivá otázka kvízu (jiný typ než 4)

0x0000	WORD	Neznámá hodnota, dost často rovna druhé hodnotě, i při nastavení na libovolnou hodnotu, klidně stejnou pro všechny otázky, jsem nezaznamenal rozdíl – pro kvíz typu 0.
0x0002	WORD	OID vedoucí na zvuk kvízové otázky
0x0004		Zde je tabulka OIDů, které označují správné odpovědi (stejný formát jako tabulka OIDů popsaná výše)

Jednotlivá otázka kvízu – typ 4

0x0000	WORD	OID vedoucí na zvuk kvízové otázky
0x0002	WORD	Neznámá hodnota, viděno 1
0x0004	WORD	Neznámá hodnota, viděno 1
0x0006	WORD	Neznámá hodnota, viděno 3
		Tabulka OIDů označujících správné odpovědi
		Tabulka OIDů, neznámý účel?
		Tabulka OIDů, zvuk správné odpovědi 1
		Tabulka OIDů, zvuk správné odpovědi 2
		Tabulka OIDů, zvuk nesprávné odpovědi 1
		Tabulka OIDů, zvuk nesprávné odpovědi 2
		Tabulka OIDů, zvuk konečného vyhodnocení pozitivního
		Tabulka OIDů, zvuk konečného vyhodnocení negativního

Šifrování MP3

MP3 soubory jsou uloženy za sebou, vždy na offsetu dělitelném 512 (0x200). Šifrování funguje takto:

Vezme se 16 bajtů uložených v hlavičce od offsetu 0x144, ke každému bajtu se přičte nejvyšší bajt header_key.
Vygeneruje se řídký klíč o velikosti 512 (0x200) bajtů a to tak, že

-na každých 16 bajtů se použijí 4 bajty klíče z hlavičky, ty se umístí na offsety 0 až 3 od čísla dělitelného 4
-tj. každých 64 bajtů se použije celý 16 bajtový klíč z hlavičky přesně jednou
-toto se zopakuje v 8 blocích
Jakým způsobem se generují offsety 0 až 3, zůstává neznámo, ale mám pohromadě tabulku, která funguje pro všechny BNL soubory – pro všechny je stejná, pravděpodobně tedy je metoda jejího odvození zabudovaná ve firmware.

Níže uvedená tabulka má pro každý bajt vstupního klíče jeden řádek, každý řádek pak říká, jaký má mít bajt klíče offset od čísla dělitelného 4 v každém z 8 bloků.
Tedy první bajt vstupního klíče bude uložen na offsetech: 0x00, 0x41, 0x81, 0xC2, 0x100, 0x141, 0x181, 0x1C2. Druhý bajt pak na offsetech 0x07, 0x47, 0x86, 0xC5, 0x105, 0x146, 0x186, 0x1C5.

```
[0,1,1,2,0,1,1,2],
[3,3,2,1,1,2,2,1],
[2,2,3,1,2,2,3,1],
[1,0,0,0,1,0,0,0],

[1,2,0,1,1,2,0,1],
[1,2,0,2,1,2,2,2],
[2,1,0,0,2,1,0,0],
[2,3,2,2,2,3,2,2],

[3,0,3,1,3,0,3,1],
[0,0,1,1,0,3,1,1],
[2,2,3,0,2,2,3,1],
[3,1,0,0,3,1,0,0],

[3,3,0,2,3,3,1,2],
[1,2,0,0,1,2,0,0],
[2,1,0,3,2,1,3,3],
[0,0,0,0,0,0,0,0]
```

Tedy

```
klíč[ blok * 0x40 + ofs_klíč * 4 + tabulka_výše[ofs_klíč][blok] = vstklíč[ofs_klíč];
```

Tento klíč se pak XORuje ke každému bajtu MP3 souboru. Vynechávají se bajty: 0x00, 0xFF, bajt shodný s bajtem klíče (protože by XOR vedl k 0x00) a bajt shodný s bajtem klíče XOR 0xFF.

Stejný postup se používá pro šifrování i dešifrování souboru.

Výchozí hodnoty

Číselné hodnoty OIDů jsou zcela jistě rozděleny do několika skupin. Původně toto rozdělení bylo odvozeno z existujících souborů, posléze, po úspěšné analýze tabulky na začátku třetí části firmware, bylo rozdělení lépe objasněno.

Kód knihy byl pozorován od 0x32A po 0xCFC, z analýzy firmware bylo posléze upřesněno na 0x2BD (701) až 0x270F (9999).

Kvízy (jejich „úvodní otázky“ a kódy natištěné na ikonách kostky) začínají vesměs od 0x64 (100). Některé kvízy ovšem vedou na neexistující OIDy zvuků úvodu – vypadá to, že pak tužka nepřehraje žádný zvuk. Podle analýzy firmware kvízy začínají na 0x64 (100) a pokračují až k 0x1F3 (499).

OIDy v OID tabulkách 0x190-0x1BD, výjimečně 0x1F4-0x221.

Většina kódů na první straně knihy začíná OID kódem 0x2AF8 (11000), 0x2AF9, 0x2AFA. Učitel začíná 0x283D, Dinosauři 0x2711 (10001). Poměrně často se dá zaznamenat, že OID kódy pro fyzické stránky knih začínají na takto zarovnaných desítkových číslech. Vzhledem ke konci rozsahu čísel knih na 9999 můžeme odhadnout, že uživatelské kódy začínají na 0x2710 (10000).

Zabudované kódy tlačítek

Ikona	**Raw kód na papíře**	**Interní kód**	**Firmware kód**
Start	Raw id knihy	0x02BD-0x270F	0x0010
Vol+	0x0015	0x0007	0x0030
Vol-	0x0016	0x0008	0x0031
Stop	0x0014	0x0006	0x0080
Porovnání	0x0531	0x0063	0x0050
MP3 mp3	0x0141	0x002E	0x0040
MP3 play	0x0143	0x002F	0x0043
MP3 pause	0x0150	0x0030	0x0042
MP3 stop	0x0151	0x0031	0x0044
MP3 prev	0x0153	0x0032	0x0045
MP3 next	0x0180	0x0033	0x0046
WAV Record 034	0xEC78	0xEE06 (609**34**)	
WAV play 031	0xEDCB	0xEF2F (612**31**)	
WAV OK any	0xEEE7	0xF03D	
Mód 1 / Otevřená kniha	0x000C	0x0004	0x0093
Mód 2 / Žárovka/Play	0x000F	0x0005	0x0094
Mód 3 / Bublina	0x0007	0x0003	0x0092
Mód 4 / Nota / Český překlad (zákl. inf.)	0x0006	0x0002	0x0091
Mód 5 / Český překlad (dialogy)	0x0005	0x0001	0x0090
Mód 6	0x25CF	0x0225	0x0095
Mód 7	0x25D5	0x0226	0x0096

Mód 8	0x25DA	0x0227	0x0097
Mód 9	0x25DF	0x0228	0x0098
Mód 10	0x25FC	0x0229	0x0099
Mód 11	0x2800	0x022A	0x009A
Mód 12	0x2819	0x022B	0x009B
Hlasitost (volume slider)	x	0x000A-0x0019	0x0020-0x002F

Existuje i řada dalších interních kódů firmware, ale zatím neznám jejich význam:
0x2000-0x201C
0x48-0x4A
0x60-0x62
0x81-0x82
0xF01-0xF16
0x70-0x76 (v novějším firmware)

Další rozsahy OIDů v hlavní konverzní funkci, nejdřív se provede toto vyhodnocení, až pak se konzultuje šifrovaná tabulka v třetí části firmware. Naprosto mi zatím uniká důvod duplikace téhož v kódu i v tabulce.

0xFE11 = 0x31 -> vol-
0xFE12 = 0x30 -> vol+
0x9 = 0x504
0xCB3A = 0x60
0xCB3B = 0x61
0xCB3C = 0x62
0xCB83 = 0x108
0x34-0x3A = 0x100-0x106
0xCB54-0xCB5A = 0x100-0x106
0x1E, 0xCB52 = 0x109
0x1F, 0xCB53 = 0x10A
0x50, 0xCB70 = 0x10B
0x51, 0xCB71 = 0x10C
0x3C-0x4F = 0x300-0x313
0xCB5C-0xCB6F = 0x314-0x327
0xEA61-0xEB8C = 0x3000-0x312B
0xF231-0xF4EB = 0x312C-0x3257
0xEB8D-0xECB8 = 0x4000-0x412B
0xF4ED-0xF7A7 = 0x412C-0x43E6
0xEDE5-0xEF10 = 0x5000-0x512B -> wav record?
0xEF11-0xF03C = 0x6000-0x612B -> wav play?
0xECB9-0xEDE4 = 0x7000
0xF03D-0xF168 = 0x7001
0x52,0xCB72 = 0xE6
0x62,0xCB82 = 0xEF
0x5D-0x61 = 0xE8-0xEC
0xCB7D-0xCB81 = 0xE8-0xEC
0x53-0x5C = 0xF0-0xF9
0xCB73-0xCB7C = 0xF0-0xF9
0xD2F1-0xE14A = 0x40C -> písničky?
0x215 = 0x501

0x216 = 0x502
0x217 = 0x404
0x218 = 0x405
0x219 = 0x406
0x21A = 0x407
0x21B = 0x408
0x21C = 0x503

Historie

Verze popisu 1.0, 2.2.2022 jindroush – první verze
Verze popisu 1.1, 21.2.2022 jindroush – přidány kódy tlačítek
Verze popisu 1.2, 5.3.2022 jindroush – přidány přesnější rozsahy kódů
Verze popisu 1.3, 7.4.2022 jindroush – přidány další rozsahy kódů