INSTANT MESSAGING NETWORK ANALYSIS AND RE-CONSTRUCTION (YMSG AND ICQ)

Jan Pluskal

Bachelor Degree Programme (3), FIT BUT

E-mail: xplusk03@stud.fit.vutbr.cz

Supervised by: Vladimír Veselý

E-mail: ivesely@fit.vutbr.cz

Abstract: This paper describes basic facts about two common instant messaging services - ICQ and Yahoo! Messenger and their communication protocols OSCAR and YMSG. Brief analysis of these proprietary protocols is provided. Two ways are shown how to decide whether the analysis is correct. The first one used implementation of experimental tool which reconstruct cached YMSG communication. The second one is NPL script which parses cached OSCAR communication and displays it in user friendly way on Microsoft network monitor [1].

Keywords: ICQ, OSCAR, Yahoo! Messenger, YMSG, instant messaging reconstruction, sniffing

1. ÚVOD

V této práci se budeme zabývat rekonstrukcí instantní komunikace klientů *Yahoo! Messenger* a *ICQ*. Celému snažení předchází analýza použitých protokolů. Pro *ICQ* je jím protokol *OSCAR* a pro *Yahoo! Messenger* je tímto protokolem *YMSG*. Samozřejmostí je znalost protokolů nižších vrstev, aby bylo možné provést korektní vypouzdření dat.

Po analýze následuje popis dvou způsobů ověření, zdali analýza byla úspěšná. Prvním z nich je vytvoření experimentálního nástroje na základě analýzy protokolu *YMSG*. Druhým je popis analyzačního skriptu pro protokol *OSCAR* v jazyce *NPL* [2].

2. ANALÝZA PROTOKOLŮ

Jelikož *OSCAR* a *YMSG* jsou proprietární protokoly, není možné k nim získat oficiální dokumentaci. Proto jako podklady sloužily informace získané z internetových fór a osobních stránek expertů zabývajících se jejich rekonstrukcí [3, 4, 5]. Relevantnost těchto informací bylo nutné ověřit v praxi, a proto byly použity jako předloha pro následnou analýzu metodou reverzního inženýrství za pomoci nástroje *WireShark*. Jediným dodatečně použitým prokazatelně relevantním zdrojem byl článek z IEEE Digital Library [7].

2.1. OSCAR

Protokol *OSCAR* je binární a skládá se z několika vrstev. Komunikace probíhá nad *TCP/IP* na portu *5190*. Na nejnižší vrstvě se nachází nízko úrovňový komunikační protokol s názvem *FLAP*. O vrstvu výše se je obvykle základní komunikační jednotka s názvem *SNAC*. Ta ve své datové části přenáší již konkrétní typy zpráv, které je důležité identifikovat pro pozdější rekonstrukci.

Komunikační protokol *FLAP* umožňuje vývoj vyšších datagramově orientovaných komunikačních vrstev. Je použit ve všech typech zpráv zasílaných mezi klienty a serverem. Počáteční *byte* nabývá hodnoty *0x2A* a je jím signalizován začátek datagramu. Ve své hlavičce obsahuje informace o *kanálu*, *sekvenčním čísle* a *velikosti* následující *datové části*.

Komunikační jednotka *SNAC* tvoří základní kámen *OSCAR* protokolu. Je přenášena na druhém kanálu *FLAP* vrstvy a ve své hlavičce obsahuje identifikaci typu následujících dat. Těmito daty jsou přenášené zprávy, které se podle určení významu dělí na rodiny a následně na podtypy.

Dílčí informace typické pro jednotlivé podtypy jsou přenášeny jako *TLV* (typ-délka-hodnota) záznamy. Jejich presence či absence blíže určuje typ zprávy.

Pro komplexnější služby jako např. *přenos souborů* může být použito několik typů spojení: *přímé propojení TCP/IP*, nebo *směrování* přes *Rendezvous point* (server s veřejnou IP adresou) a jiné. Vyjednávání o typu připojení, a servisní či signalizační informace jsou přenášeny přes protokol *OSCAR* jako speciální typ zprávy.

Podobně je tomu tak i u *video přenosu* nebo *VoIP*, kde obě služby používají protokol *SIP* a signalizace je opět přenášena pomocí speciálního typu zprávy protokolu *OSCAR*.

2.2. YMSG

YMSG je plain-textový komunikační protokol nad *TCP/IP* na portu *5050*. V pevně dané hlavičce první čtyři *byty* identifikují protokol *ASCII* konstantou „YMSG“. Následuje *číslo verze, ID výrobce, délka* a *typ zprávy* v datové části, *status* a *identifikace relace*.

Každá zpráva svým typem označuje jedinečnou akci. V datové části může (nemusí) obsahovat *TV* (typ-hodnota) záznamy. Typ každého záznamu je unikátní pro celý *YMSG* protokol, přičemž záznam jednoho typu se může objevit ve více typech zpráv. Jako oddělovač TV záznamu je použita konstanta *0xC080*.

Při hlubší analýze možností klienta *Yahoo! Messenger* bylo zjištěno, že protokol *YMSG* obstarává pouze základní funkčnost spojenou s instantní komunikací. Pokročilejší funkčnost je závislá na ostatních protokolech. Pro *přenos souborů* je použit protokol *HTTP*, který jako jediný je možné rekonstruovat, protože pro *video přenos* (*H.323*) a *VoIP* (*SIP*) je použito šifrování *SSL/TSL*.

3. EXPERIMENTÁLNÍ NÁSTROJ PRO REKONSTRUKCI YMSG

Na základě analýzy protokolu byla vybrána množina zpráv, které přenášely klíčové informace, které byly pro rekonstrukci podstatné např. *textová zpráva* mezi klientem a kontaktem, nebo *načtení listu kontaktů* a mnohé další. Pro rekonstrukci těchto zpráv byl implementován experimentální nástroj, který zároveň sloužil jako model pro pokusné ověření správnosti analýzy.

Vstupem modelu byl pro každý test předem vytvořený *PCAP* soubor se známým obsahem, který byl modelem zpracován a byl vytvořen výstup ve formátu XML souboru. Výjimku tvoří rekonstrukce přenosu souboru, kde je třeba vytvořit souvislé *TCP* toky např. za pomoci nástroje *tcpflow*. Vybrat tok obsahující požadovaný přenos a soubor z něj vypouzdřit.

Všechny provedené testy byly úspěšné a výstupy přesně odpovídaly akcím, které byly provedeny při vytváření *PCAP* souborů.

4. PARSOVÁNÍ POMOCÍ NPL

Network monitor parsing language (*NPL*) je jazyk založený na skriptování poskytující možnosti popisu organizace dat v jednotlivých protokolech. Je vyvíjen spolu s programem pro zachycení a analýzu síťové komunikace *Microsoft network monitor* [1]. Výhodou je možnost vytváření zásuvných modulů v jazyce *C/C++* a jimi řešit konstrukce, které v *NPL* nejsou možné.

Po dohodě s vedoucím práce a pro potřeby projektu *ANSA* [6] byly vytvořeny *NPL* analyzační skripty pro rekonstrukci protokolů *OSCAR* a *YMSG*.

4.1. OSCAR

Vytvoření skriptu v jazyce *NPL* se ukázalo jako velice výhodné, protože během něj bylo objeveno několik nových typů zpráv, které se v předchozích verzích protokolu nenacházely, nebo nebyly dokumentované v použitých zdrojích. Bylo také nalezeno několik chyb ve filtru, který používá *Wireshark*, pro analýzu a zobrazení zachycené komunikace.

4.2. YMSG

Během vytváření *NPL* popisu pro analýzu komunikace *YMSG* se vyskytl problém, jak získat záznamy *TV* (typ-hodnota), kde jak *typ*, tak *hodnota* jsou *ASCII* řetězce oddělená hexadecimální sekvencí *0xC080*. Samotné *NPL* obsahuje funkce vracející jak *ASCII* tak i *UTF-8* nebo *UTF-16* řetězce, ale tyto funkce fungují korektně pouze, pokud je jako ukončovač použitý řetězec, kde každý znak má *ordinální hodnotou* menší než *127*.

Pokud se nepodaří najít jiné řešení, bude nutné tento problém vyřešit vytvořením zásuvného modulu v jazyce *C/C++* obsahujícím funkci pro získání *ASCII* řetězce ukončeného libovolnou sekvencí znaků. Tato možnost by ovšem znesnadnila instalaci analyzačního skriptu.

5. ZÁVĚR

Vytvořením experimentálního nástroje pro rekonstrukci *YMSG* komunikace byla ověřena správnost analýzy tohoto protokolu. Vytvoření *NPL* popisu protokolu *OSCAR* a porovnání výsledků s výstupem programu *Wireshark* ověřilo správnost analýzy protokolu *OSCAR*.

Protože oba protokoly jsou proprietární a můžou se kdykoliv změnit a protože bez nahlédnutí do aktuálních zdrojových kódů oficiálních klientů nemůžeme nikdy potvrdit absolutní správnost, výsledkem práce je, že zatím vše funguje podle očekávání.

PODĚKOVÁNÍ

Tento příspěvek vznikl za podpory projektu MŠMT CZ.1.07/2.3.00/09.0067 TeamIT – Budování konkurenceschopných výzkumných týmů pro IT.

REFERENCE

[1] MICROSOFT. Information about Network Monitor 3 [online]. 23. 09. 2011 [cit. 2012-02-29]. Dostupné z WWW: <http://support.microsoft.com/kb/933741>.

[2] MICROSOFT. NPL – The Power Behind the Parsers [online]. 04. 10. 2006 [cit. 2012-02-29]. Dostupné z WWW: <http://blogs.technet.com/b/netmon/archive/2006/10/04/npl-_1320_-the-power-behind-the-parsers.aspx>.

[3] Yahoo v16 protocol. Adren Software [online]. [cit. 2012-02-28]. Dostupné z WWW: <http://www.adrensoftware.com/tools/yahoo_v16_protocol.php>.

[4] Yahoo's YMSG Protocol v16. Carbonize [online]. [cit. 2012-02-28]. Dostupné z WWW: <http://carbonize.co.uk/ymsg16.html>.

[5] OSCAR (ICQ v7/v8/v9) protocol documentation. Iserverd [online]. 07. 02. 2005 [cit. 2012-02-28]. Dostupné z WWW: <http://iserverd.khstu.ru/oscar/index.html>.

[6] Homepage. ANSAWiki [online]. 26. 10. 2010 [cit. 2012-02-28]. Dostupné z WWW: <https://nes.fit.vutbr.cz/ansa/pmwiki.php?n=Main.HomePage>.

[7] Jennings, R.B.; Nahum, E.M.; Olshefski, D.P.; Saha, D.; Zon-Yin Shae; Waters, C.; , "A study of Internet instant messaging and chat protocols," Network, IEEE , vol.20, no.4, pp.16-21, July-Aug. 2006 doi: 10.1109/MNET.2006.1668399 Dostupné z WWW: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1668399&isnumber=34939>