

Sem vložte zadání Vaší práce.

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
KATEDRA TEORETICKÉ INFORMATIKY



Diplomová práce

Comma-shell, interaktivní debugger shellu

Bc. Tomáš Nesrovnal

Vedoucí práce: Ing. Jan Baier

10. dubna 2017

Poděkování

Doplňte, máte-li komu a za co děkovat. V opačném případě úplně odstráňte tento příkaz.

Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval(a) samostatně a že jsem uvedl(a) veškeré použité informační zdroje v souladu s Metodickým pokynem o etické přípravě vysokoškolských závěrečných prací.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů. V souladu s ust. § 46 odst. 6 tohoto zákona tímto uděluji nevýhradní oprávnění (licenci) k užití této mojí práce, a to včetně všech počítačových programů, jež jsou její součástí či přílohou, a veškeré jejich dokumentace (dále souhrnně jen „Dílo“), a to všem osobám, které si přejí Dílo užít. Tyto osoby jsou oprávněny Dílo užít jakýmkoli způsobem, který nesnižuje hodnotu Díla, a za jakýmkoli účelem (včetně užití k výdělečným účelům). Toto oprávnění je časově, teritoriálně i množstevně neomezené. Každá osoba, která využije výše uvedenou licenci, se však zavazuje udělit ke každému dílu, které vznikne (byť jen zčásti) na základě Díla, úpravou Díla, spojením Díla s jiným dílem, zařazením Díla do díla souborného či zpracováním Díla (včetně překladu), licenci alespoň ve výše uvedeném rozsahu a zároveň zpřístupnit zdrojový kód takového díla alespoň srovnatelným způsobem a ve srovnatelném rozsahu, jako je zpřístupněn zdrojový kód Díla.

V Praze dne 10. dubna 2017

.....

České vysoké učení technické v Praze

Fakulta informačních technologií

© 2017 Tomáš Nesrovnal. Všechna práva vyhrazena.

Tato práce vznikla jako školní dílo na Českém vysokém učení technickém v Praze, Fakultě informačních technologií. Práce je chráněna právními předpisy a mezinárodními úmluvami o právu autorském a právech souvisejících s právem autorským. K jejímu užití, s výjimkou bezúplatných zákonných licencí, je nezbytný souhlas autora.

Odkaz na tuto práci

Nesrovnal, Tomáš. *Comma-shell, interaktivní debugger shellu*. Diplomová práce. Praha: České vysoké učení technické v Praze, Fakulta informačních technologií, 2017. Dostupný také z WWW: (<https://github.com/nesro/nesrotom-dip-2016>).

Abstrakt

V několika větách shrňte obsah a přínos této práce v češtině. Po přečtení abstraktu by se čtenář měl mít čtenář dost informací pro rozhodnutí, zda chce Vaši práci číst.

Klíčová slova Nahradte seznamem klíčových slov v češtině oddělených čárkou.

Abstract

Sem doplňte ekvivalent abstraktu Vaší práce v angličtině.

Keywords Nahradte seznamem klíčových slov v angličtině oddělených čárkou.

Obsah

Úvod	1
Motivace	1
1 Definice a pojmy	3
2 Historie shellu a dnešní využití	5
2.1 UNIX	5
2.2 Linux	5
2.3 Shell	5
2.4 Bash - Bourne Again Shell	6
2.5 Sh - Bourne Shell	6
2.6 Dash - Debian Almquist Shell	6
2.7 Zsh - Z Shell	6
2.8 Práce v příkazové řádce	6
2.9 Nastavení shellu	6
2.10 Terminál	6
2.11 Existující řešení	7
2.12 Bash frameworky pro psaní skriptů	7
2.13 Bash frameworky pro správu doplňků	7
2.14 Bash frameworky pro úpravu promptu	8
3 Analýza a návrh	9
3.1 Cíl práce	9
3.2 Fungování shellu	9
3.3 Debugování shellu	10
3.4 Možnosti debugování v interaktivním shellu	11
3.5 Statická analýza skriptů	11
4 Realizace	13
4.1 Nespouštění příkazů	13

4.2	Implementace debuggeru	13
4.3	Hooks	13
4.4	Bezpečný mód	14
4.5	Historie	15
4.6	Instalace	15
4.7	Testování	15
5	Testování	17
5.1	Automatizované spouštění příkazů	17
	Závěr	19
	Literatura	21
	A Seznam použitých zkratk	23
	B Obsah přiloženého CD	25

Seznam obrázků

2.1	Historie Shellu	5
-----	---------------------------	---

Úvod

Grafické uživatelské rozhraní (GUI) se jednoduše ovládá, ale ne vždy je k dispozici. To platí zejména při ovládání serverů.

Rozhraní příkazové řádky (CLI) je základní textové prostředí pro komunikaci s operačním systémem. Umožňuje spouštění programů, vkládat vstupní data a sledovat výstupní data v terminálu.

Jedním ze základních bodů UNIXové filosofie je mít jednoduché programy, které dělají pouze jednu věc, ale dělají ji dobře. To platí zejména pro základní příkazy ze sady GNU coreutils, tedy příkazy pro základní manipulaci se soubory, shellem a textem.

Tyto základní příkazy je možné řetězit a tím vytvářet užitečné jednořádkové skripty.

TODO: napsat o tom, že pro začátečníka to může být neintuitivní, musí si pamatovat spoustu příkazů. o návratových kódech, o historii příkazu a o logování

Motivace

Příkazová řádka a základní práce s ní by měly být jedna z prvních věcí, které by se uživatel měl naučit, pokud chce umět používat svůj systém na vyšší úrovni. Stejně jako v programování platí, že počítač provádí dokonale a přesně to, co mu řekneme. Ale i při sebemenší chybě v našem příkazu selže celá sekvence akcí, které se měly provést. Není těžké udělat i nějakou fatální chybu, po které není snadné uvést vše do původního stavu.

Náš projekt, Comma-shell, si dává za úkol vytvořit z příkazové řádky místo, ve kterém je těžší nějaké chyby udělat, případně před chybou varovat, nebo vysvětlit, jak danou věc udělat lépe. Důležitý cíl je také umožnit nastavit toto prostředí tak, aby bylo snadno použitelné, nebylo potřeba externích terminálů, nebo grafických aplikací a aby bylo možné snadno omezit repetitivní výstupy programu. Comma-shell se dělí na tři části, každá pomáhá řešit jiný problém.

Comma-shell hooks, nebo-li části kódu, které lze vykonat před, nebo po vykonání spuštěného příkazu a dokonce zabránit jeho spuštění umožňují ochránit uživatele před vykonáním příkazu, který je v například nějakém smyslu nesprávný, špatný, nebo je v nepořádku. Pokud tedy napíše příkaz, který obsahuje nějaký překlep, nebo chybu, která by způsobila neočekávané chování, může být uživatel varován, než bude takový příkaz proveden. Nejde jen o ochranu uživatele před špatnými příkazy. Díky informacím o spouštěném příkazu a následně o jeho výsledku lze psát různé skripty, které by jinak bylo složité vytvořit.

Comma-shell debugger, tedy ladící program pro příkazy spouštěné z interaktivní příkazové řádky. Pokud uživatel píše složitější program, obsahující například smyčku, nebo roury a příkaz nefunguje tak, jak uživatel očekává, tak nejjednodušší je rozdělit si příkaz na podpříkazy a ujistit se, že každý příkaz dělá to, co od něj uživatel očekává. Comma-shell debugger by měl tento postup automatizovat.

Poslední částí jsou takzvané bezpečné příkazy. Pokud uživatel vykoná příkaz, jehož efekt byl jiný než očekával, chce spustit opačný příkaz, aby napravit svůj omyl, nebo má často mnohem větší problém, pokud šlo například o nesprávné použití programu pro mazání souborů. Cílem bezpečných příkazů je tedy zaprvé umožnit vrátit provedené změny, ale i informovat uživatele o tom, co se stane. Pokud uživatel používá ke specifikování souborů, které mají být cílem nějakého příkazu, je častý omyl vybrat nějaké soubory nechtěně. Nebo například provést příkaz ze špatného adresáře.

Definice a pojmy

Příkazová řádka, anglicky Command Line Interface, je uživatelské rozhraní ovládané příkazy. Příkazy spuštěné z příkazové řádky vypisují výstup také na příkazovou řádku a další interakce se spuštěným programem probíhá zase zadáváním příkazů. Na rozdíl od textového nebo grafického rozhraní uživatel nemůže nijak zasahovat do již vypsané či vykreslené části.

TODO: obrazek prikazove radky, tui a gui

Shellem se obecně myslí uživatelské prostředí pro využívání funkcí jádra operačního systému. V této práci budeme slovem shell označovat právě interpret příkazů v příkazové řádce. Tedy program, ovládaný příkazy, který umí spouštět ostatní programy a zajišťovat jejich vstup a zobrazovat jejich výstup.

Prompt je krátký řetězec znaků, který se objeví na začátku řádky, do které se bude vepisovat příkaz v příkazové řádce. Je používán k zobrazování důležitých informací, jako například, ve kterém adresáři se uživatel nachází či jaké je jméno počítače, na kterém se nachází.

GNU je projekt zabývající se tvorbou a propagací svobodného software. Cílem je vytvořit kompletní svobodný operační systém unixového typu. Protože jádro operačního systému od GNU s názvem Hurd nebylo tak úspěšné, začalo se používat jádro Linux. Kombinace operačního systému GNU a jádra Linux se označuje jako GNU/Linux. GNU vytvořilo i svůj shell, Bash. Sada základních příkazů, jejichž existence je předpokládána v nějaké formě na každém operačním systému, se jmenuje GNU coreutils. Patří do ní příkazy jako ls, cp, rm, atd. Je důležité si uvědomit, že příkazy jako cd, echo jsou vestavěné příkazy shellu Bash.

Terminál byl z historického hlediska elektronické zařízení k základní komunikaci s počítačem. Dnes se v operačních systémech spouští terminálové emulátory, ve kterých lze spustit různé programy. Nejčastěji je to právě shell.

Historie shellu a dnešní využití

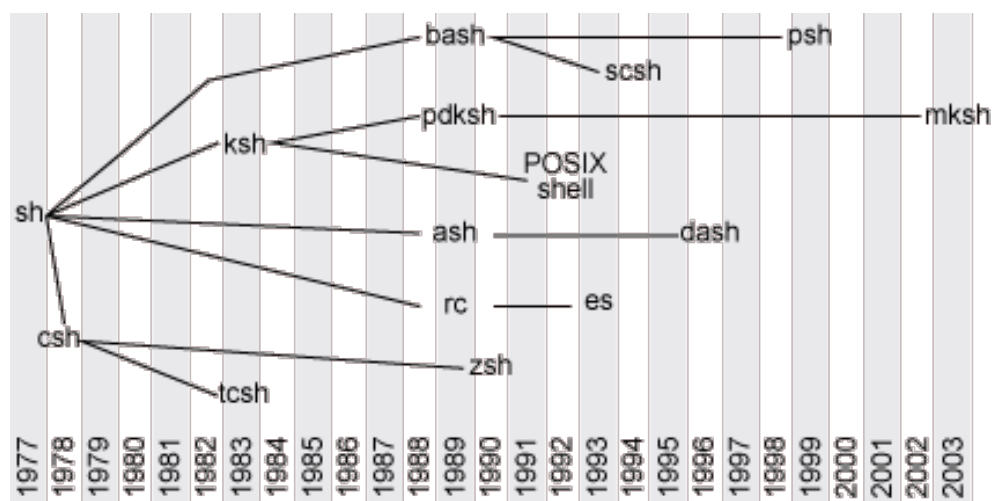
Napsat o tom, že bash je dneska nejcastejsi, protoze linux je vlastne GNU/Linux a bash je GNU bash. Zsh je lepsi snad uplne ve vsech. Sh, resp Dash je rychlejsi.

2.1 UNIX

todo: historie unixu

2.2 Linux

2.3 Shell



Obrázek 2.1: Historie Shellu

V latexových komentářích jsou nějaké odkazy o historii.

2.4 Bash - Bourne Again Shell

https://cs.wikipedia.org/wiki/Debian_Almquist_shell <https://wiki.archlinux.org/index.php/Dash>

2.5 Sh - Bourne Shell

todo

2.6 Dash - Debian Almquist Shell

todo

2.7 Zsh - Z Shell

todo

<http://zsh.sourceforge.net/FAQ/>

2.8 Práce v příkazové řádce

todo: tady bych chtěl shrnout obecně to, že se píšou příkazy jako text a uživatel vše vidí jako text. co z toho plyne za následky a jak je snadné udělat chybu

2.9 Nastavení shellu

GNU/Linuxové distribuce, které nabízejí předpřipravené prostředí, mají pro výchozí shell bash připravené takzvané rc soubory, které nějakým způsobem upravují běžící interaktivní shell. Jednou z nejviditelnějších změn je nastavení proměnné PS1, o které bude řeč v pozdější kapitole, která určuje styl promptu. Ve výchozím nastavení bashe, tedy po spuštění bashe bez načtení rc souborů (`bash -norc`), se v PS1 zobrazuje pouze název shellu, jeho verze a zdali je uživatel root. Po načtení výchozích rc souborů se v PS1 ukazují informace jako je například jméno uživatele, jméno počítače a co je velmi důležité, jméno aktuálního adresáře.

Tyto informace jdou snadno získat použitím některých základních příkazů, například `whoami`, `hostname`, `pwd`. Protože tyto informace jsou důležité a potřebujeme je vědět pořád, chceme je mít pořád na očích.

2.10 Terminál

todo: asi by se hodilo napsat i něco o těch programech, ve kterých shell běží

Napsat o tom, že cílem práce je usnadnit práci v příkazové řádce a sepsat základní funkcionalitu debuggeru.

2.11 Existující řešení

Mezi seznamem různých zajímavých řešení awesome (github.com/sindresorhus/awesome) existuje i podsekcce awesome-shell (github.com/alebcay/awesome-shell), ve které lze nalézt spoustu užitečných nástrojů pro práci s příkazovou řádkou, nebo psaním skriptů v bashi.

Spousta těchto nástrojů je nad rámec této práce.

2.12 Bash frameworky pro psaní skriptů

Existuje mnoho projektů, jejichž cílem je vytvořit framework v bashi, který má nějakým způsobem zjednodušit vytváření, především větších, skriptů v bashi. Spousta dnešních vývojářů je zvyklá na objektově orientovaný přístup k programování a je tedy pro těžké vytvořit větší program, který by zůstal přehledný.

2.12.1 Bash OO framework

Bash OO framework (github.com/niieani/bash-oo-framework) je framework napasný v bashi, který umožňuje vytvářet třídy, vyjimky a testy. Jeho cílem je vytvořit prostředí pro psaní skriptu, kde se bude snadněji psát čitelný kód, bez částí, které se opakují.

2.13 Bash frameworky pro správu doplňků

Existuje spousta věcí, které uživatel příkazové řádky potřebuje občas řešit. Řešení pro daný problém je víc. Buď si uživatel vyřeší problém sám, nebo bude hledat řešení na internetu. Shellové frameworky, které umějí i spravovat doplňky, mohou mít řešení pro daný problém. Instalace a použití pak bude velmi jednoduché, intuitivní a bude zde nějaká záruka o funkčnosti a kvalitě.

Přestože je tato práce převážně o shellu bash, zsh je v tomto mnohem rozšířenější.

2.13.1 Oh My Zsh

Oh My Zsh ([ohmyz.sh](https://github.com/ohmyzsh)) je framework pro Zsh.

2.13.2 Bash-it

Bash-it (<http://github.com/Bash-it/bash-it>)
(itsfoss.com/bash-it-terminal-tool/)

2.14 Bash frameworky pro úpravu promptu

2.14.1 Liquid Prompt

Projekt liquidprompt github.com/nojhan/liquidprompt je adaptivní prompt v interaktivním bashi.

2.14.2 Sexy Bash Prompt

Projekt sexy-bash-prompt <https://github.com/twolfson/sexy-bash-prompt> je další používaný prompt v interaktivním bashi.

Analýza a návrh

3.1 Cíl práce

todo

3.2 Fungování shellu

3.2.1 Životní cyklus příkazu

todo: popsát, co všechno se stane, od napsání příkazu až po jeho vykonání

3.2.2 Gramatika shellu

Soubor s gramatikou bash, `parse.y`, má přes 6000 řádek. Chtěl bych zde napsat zjednodušenou gramatiku, která by se dala snadno pochopit (ono to zatím tak složité není).

Popsat, jakým způsobem parsuje gramatiku BASH (yacc) a jakým to dělají parsery `bashlex` a `bashast`.

3.2.2.1 Projekty parsující shell

<https://github.com/bemeurer/beautysh/blob/master/beautysh/beautysh.py>

3.2.2.2 Bashlex

<https://github.com/idank/bashlex/blob/master/bashlex/parser.py>

3.2.2.3 Bashast

<https://github.com/neloe/libbash/blob/master/bashast/bashast.g>

3.2.3 Spouštění příkazů

Popsat základní principy jak funguje shell. Popsat procesy v unixu, fork, exec, co všechno se musí stát, aby shell mohl spustit příkaz.

3.2.4 Struktura BASHe

todo: Zdrojový kód je rozdělen do souboru, možná by bylo dobře popsát popsát co který soubor dělá, aby si čtenář udělal alespoň trochu obrázek.

3.3 Debugování shellu

tohle je jeden bod zadání: "Proveďte řešení existujících nástrojů pro statickou analýzu, krokování a hledání chyb v BASH skriptech."

3.3.1 Debugování Bashe

todo: návod na debugování bashe přímo pomocí gdb. (tohle nakonec pro moji práci nebylo potřeba)

3.3.2 Interní nástroje

todo: popsát to, jaké nástroje má v sobě bash zabudované v základu

3.3.2.1 Debugovací mód BASHe (jak funguje shopt s extdebug)

shopt s extdebug

3.3.2.2 set x, u, v, e

příklady do skriptu

3.3.2.3 PS0, PS4

PS0 bude v novém bashi, my ji proto nebudeme používat, PS4 se vypisuje při debugování. todo: ps0 se nakonec nepovedlo

3.3.3 Externí nástroje

3.3.4 BASH Debugger

todo: popsát jak funguje, co všechno umí, nějaké příklady <http://bashdb.sourceforge.net/>

3.3.5 BashEclipse

BashEclipse je plugin do Eclipse, který umí krokovat v GUI Eclipse.

<http://unix.stackexchange.com/questions/131491/is-there-a-gui-debugger-for-shell-scripts> <https://sourceforge.net/projects/shelled/>
<https://sourceforge.net/projects/basheclipse/>

3.4 Možnosti debugování v interaktivním shellu

Nebyl nalezen žádný nástroj pro debugování interaktivního shellu. todo

3.4.1 Definice debugování interaktivního shellu

todo: nema cenu debugovat jednoduche prikazy, ale slozite ano. chceme umet rozkrokovat pipy a subshelly (tedy to co nam to dela ted)

3.4.2 GNU Readline

GNU Readline umožňuje přemapovat enter tak, abysme mohli spustit příkaz v naší definované funkci. Problémem je, že takto upravený příkaz se uloží do historie. Dalším problémem jsou víceřádkové příkazy, tedy takové, pro jejichž napsání musíme několikrát zmáchnout enter. TODO: ukázka.

3.4.3 Napsání nového REPLu

Zprovoznění základní funkcionality by bylo snadné, vzhledem ke komplexnosti BASHe však téměř nemožné mít stejné chování jako v BASHi.

3.4.4 DEBUG trap

Současné řešení. Při zapnutém extdebug je možné příkazy nepustit a jen evalovat poslední příkaz z historie. TODO: je potřeba popsat základní chování historie (např. mezera na začátku příkazu, atd.)

3.5 Statická analýza skriptů

bash má validaci syntaxe: `bash -n`, ale to nezjistí skoro žádné chyby.

3.5.1 Check Bashisms

<http://checkbashisms.sourceforge.net/> Některé skripty cheme kvůli rychlosti spouštět v Bourne-Shellu. Protože ten je pouze podmnožina bashu, tento nástroj najde výskyty syntaxe, která je podporovaná pouze v bashi a v Bourne-Shellu nemusí fungovat správně.

3.5.2 Explain Shell

Tato analýza nehledá chyby, ale vyparsuje vložený kód a vyhledá příslušnou dokumentaci v manuálových stránkách. Využívá knihovny bashlex, kterou využijeme v našem debuggeru. <http://www.explainshell.com/>

3.5.3 ShellCheck

todo: tady bych se mohl rozepsat víc. o tom, že to je napsané v haskellu, o tom že existuje databáze věcí, které se ve skriptech hledají a že ke každým je wiki..

Realizace

v zadání je: Navrhněte a implementujte nástroj, který umožní psát uživatelské skripty pro analýzu příkazů a ovlivňování jejich spouštění a vykonávání. = to jsou hooky

Nástroj musí umožňovat krokovat složitější skripty po jednotlivých příkazech. = to je debugger

Pro analýzu spouštěných skriptů využijte vhodný nástroj z řešeršní části. = to je shellcheck

4.1 Nespouštění příkazů

Pro zabránění spouštění používáme DEBUG trap. todo: popsat problémy a řešení

4.2 Implementace debuggeru

todo: napsat o tom, že část je napsána v pythonu a že se celá věc spouští v pre-hooku, takže to je přesně ten čas, kdy se může příkaz debuggovat a originální příkaz se nespustí.

popsat parsování v bashlex a jak se celá věc předává do bashu

4.3 Hooks

Aby byl kód přehledný, kromě bezpečných příkazů je funkcionality rozdělena do hooků, nebo-li modulů, které obsahují kód, který je spuštěn před, nebo i po vykonání příkazu. Kód vykonaný před příkazem může rozhodnout, zda-li má dojít k zabránění vykonání příkazu.

4.3.1 Implementace hooků

todo: napsat o tom, že se hooky pouští automaticky před i po a že je snadné vytvořit vlastní hook

4.3.2 Hooky před vykonáním příkazu

todo: napsat co všechno můžu dělat

4.3.3 Hooky po vykonání příkazu

todo: napsat co všechno můžu dělat

4.3.4 Předvytvořené hooky

todo: napsat o tom, že

4.3.5 Shellcheck hook

todo: tohle je nejlepší věc, kterou celé commash umí. tady by to možná chtělo ukázat případy, kdy se to hodí. taky popsat to, že některé věci jdou snadno zakázat

4.3.6 Bashlex hook

todo: tohle je bashovská část debuggeru

4.3.7 Explain RC hook

4.3.8 Notfound hook

4.4 Bezpečný mód

Bezpečný mód umožňuje dvě základní věci. Tou jednodušší je pouze vypnutí efektu příkazu, který má nějaké destruktivní následky. Složitější varianta dovoluje vrácení do stavu před vykonáním příkazu.

todo: napsat o tom, že po zkoušení jsem dosel k závěru, že je lepší uživateli nejprve ukázat co se děje, co se stane a jaké má možnosti. častokrát už

4.4.1 Bezpečné rm

todo: napsat o tom, jak to je vyřešené a proč. napsat, že `rm -i` i `rm -I` je k ničemu a proč je naše verze lepší popsat, jak funguje freedesktop.org trash a že ukládáme seznam souborů, které byly smazány jedním `rm`, aby se celý spuštěný příkaz `rm` vrátil

4.5 Historie

Popsat jak jsem vyresil ukladani historie prikazu, ukladani vystupu, navratove kody, jak vracet nasledky prikazu do puvodniho stavu. todo: historie taky neni v zadani a moc jsem to nedoresil. teoreticky by na to sel napsat pre a post hook, ktery zaznamena vsechno potrebné a zapise to nekam do souboru. todo: slo by snadno udelat: podobna historie co je ted, jen se ulozi odkud se prikaz spustil

4.6 Instalace

todo: odkaz na github, vypsát sem to je i tam, tedy: tvorba virtualniho stroje a instalace popsát, ze se predopklada instalace do `/.commash`

4.7 Testování

Testování

5.0.1 Testování shell skriptů

Popsat jak fungují nektě testovací frameworky. shunit2, roundup

5.1 Automatizované spouštění příkazů

Popsat jak se dají automaticky spouštět příkazy. At už lokálně, nebo vzdalene. Popsat jak rekonstrukci z typescriptu, tak třeba Tcl, Expect.

5.1.1 Expect

todo: jak se pracuje s expectem

Pro testování byl zvolen skriptovací jazyk Expect, který je rozšířením jazyka Tcl. Lze tedy velice snadno otestovat chování Comma-shellu a jeho částí pro různé vstupy.

Testy se skládají ze vstupů a z řetězců, které se očekávají na výstupu. Přes Expect se potom vytvoří nová instance bashe, ve které se automaticky načte Comma-shell a postupně se posílají vstupy a sledují se výstupy pro očekávané řetězce.

Testy jsou rozdělené na menší části, kde každá testuje jednu konkrétní věc. Přes skript lze pustit sérii všech testů.

5.1.2 Testování Debuggeru

5.1.3 Testování Hooks

5.1.4 Testování Bezpečných příkazů

5.1.5 Měření paměťových nároků

Závěr

todo: urcite bych napsal. ze nikdo nic podobneho neudelal a ohlasy byly celkem pozitivni. pokud se toho chytne alespon par lidi a budou hlasit chyby, co je potreba vylepsit, nebo dokonce i prispivat, mohl by z toho byt jednou opravdu vyborny pomocnik

sem napište závěr Vaší práce

Literatura

Seznam použitých zkratk

GUI Graphical user interface

XML Extensible markup language

Obsah přiloženého CD

	readme.txt.....	stručný popis obsahu CD
	exe	adresář se spustitelnou formou implementace
	src	
	impl.....	zdrojové kódy implementace
	thesis	zdrojová forma práce ve formátu L ^A T _E X
	text	text práce
	thesis.pdf	text práce ve formátu PDF
	thesis.ps	text práce ve formátu PS