Posudek oponenta závěrečné práce

České vysoké učení technické v Praze **Fakulta informačních technologií**

Student: **Jakub Samek**
Oponent práce: **Ing. Josef Hlaváč, Ph.D.**
Název práce: **Virtuální GPU cluster**

Typ práce: *BI*
Datum vytvoření hodnocení: 26.2.2015

Hodnotící kritérium:	*Způsob hodnocení - následující škálou 1 až 5:*
1. Náročnost a další komentář k zadání	*1=mimořádně náročné zadání,* ***2=náročnější zadání,*** *3=průměrně náročné zadání,* *4=lehčí, ale ještě dostatečně náročné zadání,* *5=nedostatečně náročné zadání*

Popis kritéria:
Podrobněji charakterizujte diplomovou (bakalářskou) práci a její případné návaznosti na předchozí nebo běžící projekty. Dále posuďte, čím je zadání této ZP náročné. (U obtížnější ZP lze dále tolerovat některé nedostatky, které by u ZP standardní obtížnosti tolerovány nebyly; a naopak u jednoduché ZP mohou být zjištěné nedostatky hodnoceny přísněji.)

Komentář: Práce se zabývá problematikou virtuálních clusterů GPU a jejich využití k prolamování hesel hrubou silou. Jde o téma velmi aktuální, neboť bezpečnost počítačových systémů je jedním z palčivých problémů současnosti. Autor po teoretickém úvodu uvádí podrobný popis realizace GPU clusteru v počítačových učebnách FIT ČVUT. Vzhledem k tomu, že autor vedle využití stávajících nástrojů vyvinul i nástroj vlastní, hodnotím zadání jako náročnější.

Hodnotící kritérium:	*Způsob hodnocení - následující škálou 1 až 4:*
2. Splnění zadání	***1=zadání splněno,*** *2=zadání splněno s menšími výhradami,* *3=zadání splněno s většími výhradami,* *4=zadání nesplněno*

Popis kritéria:
Posuďte, zda předložená ZP splňuje zadání. V komentáři uveďte body zadání, které nebyly zcela splněny, případně rozšíření ZP oproti původnímu zadání. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.

Komentář: Zadání považuji za splněné.

Hodnotící kritérium:	*Způsob hodnocení - následující škálou 1 až 4:*
3. Rozsah písemné zprávy	*1=splňuje požadavky,* ***2=splňuje požadavky s menšími výhradami,*** *3=splňuje požadavky s většími výhradami,* *4=nesplňuje požadavky*

Popis kritéria:
Porovnejte rozsah předložené písemné zprávy s požadovaným rozsahem, viz Směrnice děkana č. 9/2011, článek 3. Pro hodnocení ZP je také důležité, zda všechny části písemné zprávy jsou informačně bohaté a pro práci nezbytné. Text ZP by neměl obsahovat zbytečné části.

Komentář: Písemná zpráva je poměrně rozsáhlá (62 stran + krátké přílohy). Autor místy zabíhá do zbytečných podrobností (např. IP rozsahy -- viz dále). Část kapitoly 3.1.6, zejména popis konfigurace a příkazů, by bylo vhodnější uvést v příloze, ideálně v podobě manuálové stránky.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
4. Věcná a logická úroveň práce	*80 (B)*

Popis kritéria:
Posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti. Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře.

Komentář: Práce je po věcné a logické stránce v zásadě v pořádku. Mám však následující výhrady:
-- Autor není konzistentní v používané terminologii. Hned v první větě v kapitole 1.1 zavádí termín "ověřování totožnosti" a synonymum "autentizace", ale v následujícím textu často používá jako další synonymum "autentifikace".
-- Autor se dopouští určitých nepřesností při používání pojmu "entropie". V definici v kapitole 1.1.2 uvádí, že "pro popsání stavu, ve kterém se systém nachází, je potřeba přesně tolik bitů informace, jaká je jeho entropie". Bylo by vhodnější nahradit slovo "přesně" slovem "nejméně", neboť případná redundance v popisu nějakého stavu nezvyšuje jeho entropii. Dále na konci kapitoly 1.2.3 autor tvrdí, že "heslo dnesjekrasne1 nelze považovat za bezpečné, i když má velkou entropii" -- v tomto tvrzení vidím logický rozpor, viz otázka k obhajobě.
-- Polemizoval bych s hodnotou 10 TH/s jako "postavitelným maximem" (kapitola 1.2.3). Autor zahrnuje do srovnání zařízení s cenou v řádu stovek tisíc eur, což je částka, za kterou lze navrhnout a vyrobit menší sérii obvodů ASIC.
-- Nevím, proč autor v kapitolách 2.1.1 až 2.1.3 uvádí IP adresy školních počítačů. Pro účely písemné zprávy je to údaj zbytečný, navíc jeho zveřejnění může napomoci případným hackerům (byť lze namítnout, že "zabezpečení" založené na "utajení" IP adresy vnitřní sítě snad ani nelze nazývat zabezpečením). Podobnou výhradu mám k podrobnostem o privilegovaném uživateli na konci kapitoly 3.

Na druhou stranu pozitivně hodnotím tyto aspekty práce:
-- Autor při hodnocení výkonnosti nezapomíná na spotřebu energie, což je důležitý, ale často opomíjený parametr.
-- Autor informoval autora použitého softwaru o nalezených chybách (str. 24).

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
5. Formální úroveň práce	*75 (C)*

Popis kritéria:
Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 9/2011, článek 3.

Komentář: Autor poměrně často chybuje v psaní čárek, v práci se vyskytují překlepy. Některé tabulky (3.1, 3.2) obsahují podivně zarovnaný text. Tyto formální chyby však nebrání porozumění textu.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
6. Práce se zdroji	*90 (A)*

Popis kritéria:
Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení ZP. Charakterizujte výběr studijních pramenů. Posuďte, zda student využil všechny relevantní zdroje nebo zda se pokoušel řešit již vyřešené problémy. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Komentář: Kladně hodnotím, že ráce cituje velké množství zdrojů. Všechny zdroje jsou internetové, což nepovažuji za problém. Řadu bibliografických záznamů by však bylo vhodné doplnit minimálně o jméno autora (je-li známo). Postrádám zdroj údaje o rychlosti růstu výkonu GPU a CPU (str. 1).

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
7. Hodnocení výsledků, publikační výstupy a ocenění	*90 (A)*

Popis kritéria:
Vyjádřete se k úrovni dosažených hlavních výsledků ZP, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, apod. Případně také zhodnoťte, zda software nebo zdrojové texty, které nevytvořil sám student, byly v ZP použity v souladu s licenčními podmínkami a autorským právem. Popište případnou publikační činnost a získaná ocenění související s řešením této ZP.

Komentář: Autor úspěšně vytvořil virtuální GPU cluster z počítačů dostupných na fakultě. Rovněž ukázal, že vytvořený cluster je reálně použitelný k zamýšlenému účelu, byť jeho výkonnost již zaostává za clustery z modernějších GPU. V rámci práce rovněž vytvořil vlastní software k automatizaci celého procesu.

Hodnotící kritérium:	*Způsob hodnocení - nehodnotí se*
8. Komentář o využitelnosti výsledků	

Popis kritéria:
Uveďte, zda hlavní výsledky ZP rozšiřují již publikované známé výsledky a/nebo přinášející zcela nové poznatky. Uveďte možnosti využití výsledků ZP v praxi.

Komentář: Viz bod 7 -- výsledky práce jsou využitelné. Oceňuji, že autor svůj software dal veřejně k dispozici na standardní platformě (github) a pod veřejnou licencí GNU GPL.

Hodnotící kritérium:	*Způsob hodnocení - nehodnotí se*
9. Otázky k obhajobě	

Popis kritéria:
Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).

Otázky:
-- Spočítejte entropii hesla tvořeného zřetězením tří náhodně zvolených slovníkových slov a jedné náhodně zvolené dekadické číslice (např. "dnesjekrasne1"). Jako slovník předpokládejte a) seznam 3000 nejčastěji používaných slov, b) Slovník spisovné češtiny pro školu a veřejnost z r. 2005, který obsahuje 45366 hesel (odvozeniny a ohýbání zanedbejte). Srovnejte s entropií hesla složeného z 13 náhodně vybraných znaků z množiny {a-z, 0-9}.
-- Proč jste v kapitole 1.2.2.3 jako zástupce FPGA zmínil zařízení, k němuž nemáte přístup a máte o něm jen omezené informace, a nikoli například zařízení COPACOBANA, které je k dispozici na FIT ČVUT?
-- Na internetu lze nalézt srovnání výkonnosti různých zařízení (včetně GPU, FPGA apod.) při těžení kryptoměn, což je v podstatě také počítání hashe; například na adrese https://en.bitcoin.it/wiki/Mining_hardware_comparison. Bylo by možné tyto výkonnosti porovnat s Vašimi výsledky (zejména kapitolou 4.3 a 4.5), a pokud ano, jsou s Vašimi výsledky v souladu?

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
10. Celkové hodnocení	*85 (B)*

Popis kritéria:
Shrňte stránky ZP studenta, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení **nemusí** být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích 1 až 9.

Text hodnocení: Zadání práce bylo splněno. Práci považuji za zdařilou a výsledky relevantní. Vzhledem k menším výhradám navrhuji hodnocení B.

Podpis oponenta práce: