Posudek *vedoucího* diplomové práce

Jméno a příjmení autora posudku: David Hauzar

Jméno a příjmení autora práce: Pavel Baštecký

Název práce: Data Modeling for Static Analysis of Web Applications

---

Vlastní text

Hlavním cílem diplomové práce bylo nalézt a vyřešit nedostatky existujícího paměťového modelu frameworku pro statickou analýzu PHP aplikací Weverca. Tyto nedostatky byly: 1) Nekorektnost některých aspektů paměťového modelu. Například nekorektní řešení operace merge dvou kontextů s odlišnými úrovněmi stacku nebo nekorektní inicializace nově definovaných proměnných ve druhé fázi analýzy. Dále 2) neefektivita paměťového modelu. Kvůli vysoké časové i paměťové náročnosti analyzátoru nebylo možné analyzovat ani relativně malé aplikace. Poslední nedostatek byl 3) obtížná modifikace a nepřehlednost paměťového modelu. Původní paměťový model byl monolitický, strukturální informace (informace o vztazích mezi proměnnými a informace o struktuře polí a objektů), informace o nestrukturovaných hodnotách proměnných a algoritmy nad paměťovým modelem se vzájemně prolínaly bez jasně definovaného rozhraní. Bez znalosti celého paměťového modelu bylo velmi obtížné hledat v paměťovém modelu chyby a optimalizovat jeho části. Z paměťového modelu se tak postupně pro ostatní vývojáře frameworku stávala„černá skříňka“. Dalším cílem diplomové práce bylo zjistit, jaký vliv budou mít jednotlivé optimalizace paměťového modelu na jeho výkon.

Všech těchto cílů se podařilo dosáhnout. Známé nekorektnosti paměťového modelu byly odstraněny. Výsledný paměťový model je zcela unikátní v tom, že modeluje sémantiku aliasů, polí, objektů, včetně sémantiky dynamického přístupu k proměnným, indexům polí a atributům objektů a vytváření implicitních objektů a polí. Bylo dosaženo extrémního zrychlení paměťového modelu a snížení paměťové náročnosti. Už při analýze kódu o velikosti v řádu stovek SLOC je zlepšení téměř dvacetinásobné. Framework je nyní schopný analyzovat kód, který byl dříve daleko za jeho možnostmi. Monolitický paměťový model byl rozdělen do jasně definovaných částí a je snadné změnit implementaci struktury paměťového modelu a jednotlivých algoritmů paměťového modelu a lze snadno měřit, jaké mají provedené změny vliv na výkon paměťového modelu.

Ač samotný text práce není dokonalý, obsahuje gramatické a stylistických chyby a některé části textu jsou obtížněji srozumitelné, oceňuji, že se podařilo dobře popsat základní principy paměťového modelu a implementované optimalizace. Tento popis je navíc doplněný rozsáhlou dokumentací na úrovni zdrojového kódu. V diplomové práci chybí sekce „Related work“, která by uvedla přehled optimalizací použitých v obdobných paměťových modelech. Vytvořit takový přehled by ale bylo velmi obtížné. Nejsem si vědom žádného popisu optimalizací v paměťovém modelu pro statickou analýzu jazyka s podobně komplexní sémantikou, jako je PHP. Existuje několik implementací paměťových modelů pro statickou analýzu jazyků PHP a JavaScript, ovšem pouze s popisem sémantiky těchto paměťových modelů. Srovnání sémantiky paměťového modelu implementovaného v rámci této práce a ostatních paměťových modelů je možné najít ve společné publikaci [1] a dále v publikaci [3].

Pavel Baštecký prokázal nevídanou schopnost soustředěné práce, invence, ale i kritického přístupu ke své práci a schopnost spolupracovat. V neposlední řadě si cením Pavlova zaujetí pro práci, kterou dělal. Výsledky, kterých dosáhl, zdaleka předčily má očekávání.

O kvalitě práce vypovídá také to, že vedla k publikaci [1] na mezinárodním workshopu ESSS 2014 a významným způsobem přispěla k publikacím frameworku Weverca [2] a [3] na konferencích SEFM 2014 a ECOOP 2015.

[1] Hauzar, David, Kofroň, Jan and Baštecký, Pavel. *Data-flow Analysis of Programs with Associative Arrays*.
In: Proceedings of the International Workshop on Engineering Safety and Security Systems (ESSS'14), Singapore, EPTCS, 2014

[2] Hauzar, David and Kofroň, Jan. *WeVerca: Web Applications Verification for PHP*.
In: Proceedings of the Software Engineering and Formal Methods - 12th International Conference (SEFM'14), Grenoble, LNCS, 2014

[3] Hauzar, David and Kofroň, Jan. *Framework for Static Analysis of PHP Applications*.
In: Proceedings of the 29th European Conference on Object-Oriented Programming, (ECOOP'15), Prague, LIPIcs, 2015

Doporučení k obhajobě:

Z výše uvedených důvodů práci *doporučuji* k obhajobě.

Vynikající práce vhodná pro soutěž studentských prací	ANO ☒

Seznam soutěží studentských prací, viz http://www.mff.cuni.cz/studium/bcmgr/prace/

Pokud jste výše zaškrtli ANO, zdůvodněte prosím svůj návrh, případně uveďte konkrétní soutěž, pro kterou je práce vhodná (rámeček lze nechat prázdný, pokud za dostatečné zdůvodnění považujete text posudku):

Ač samotný text práce není dokonalý, práci považuji za výjimečnou kvůli výsledkům, kterých bylo při její řešení dosaženo. Výsledný paměťový model je v daném kontextu zcela unikátní a práce přispěla k publikacím na mezinárodním workshopu a konferencích.

V Gif-sur-Yvette dne 28. 8. 2015

Podpis:**

** nehodící se škrtněte (vymažte)*
*** do SISu vkládejte formulář nepodepsaný (ve formátu PDF), podpis je potřeba doplnit až na vytištěný posudek.*