



Posudek oponenta dizertační práce

Ing. Lubomír Bulej

Connector-based Performance Data Collection for Component Applications

Doktorská dizertační práce Ing. Lubomíra Buleje je věnována problematice výkonového monitorování (Performance Monitoring) aplikací budovaných ze softwareových komponent. Téma, které si doktorand pro svou práci zvolil, je velice aktuální, využití nezávisle vytvářených a dobře prověřených komponent v nových aplikacích dovoluje nejen podstatně zvýšit efektivitu tvorby software, ale současně zajišťuje i vysokou funkční spolehlivost výsledného řešení při současném zkrácení doby vývoje.

Problematika výkonového sledování aplikací opírajících se o middleware prostředky, stejně jako studium metod pro tvorbu komponentních aplikací, jsou nosné směry výzkumných skupin na KSI MFF. Architektura monitorovacího systému navrhovaná v práci byla a je součástí mezinárodní spolupráce těchto skupin na projektech SOFA a CoCoME.

Za velice významnou považují skutečnost, že úspěšná budoucí implementace studovaného přístupu k výkonovému monitoringu může přispět ke zvýšení efektivitu vývoje aplikací stavěných ze softwareových komponent. Navrhované řešení, podrobně popsané v kapitolách 5 a 6, může být bezprostředním základem nástrojů pro běžnou programátorskou praxi.

Myšlenku realizace monitorovacího systému opírajícího se o informace získatelné na rozhraních komponent, v konektorech, doktorand rozvinul do podrobného návrhu architektury monitorovacího systému. Popsaný návrh dává dobrou představu o řešení monitorovacích funkcí zaměřených na časové a výkonové charakteristiky. Problematika zdrojů dat pro monitorování systémových nároků (paměť, vlákna a procesy) zůstává poněkud v pozadí, dostupným údajům doktorand věnuje jen krátkou zmínku v kapitole 4.2 a jedná se o monitorování na úrovni operačního systému. Monitorování systémových nároků jednotlivých komponent aplikace nebo systémových zdrojů využívaných použitým middleware by mohlo být podobně zajímavé jako monitorování časových aspektů a indikátorů operačního systému, vyžádalo by si však zřejmě další podporu v kódu middleware a/nebo v kódu vlastních komponent.

Text práce se opírá o metody popisu standardizované v oblasti softwareového inženýrství a zahrnované pod označením UML - hierarchie objektů, komunikační schémata. Takový popis dovoluje doktorandovi popsat architekturu navrhovaného systému s dostatečnou podrobností a může být dobrou podporou budoucího systémového návrhu a implementace.

Za určitou slabinu textu práce považují chybějící popis vhodně zvoleného experimentu - příkladu ilustrujícího výsledek navrhovaného řešení. I když popis architektury systému, jdoucí hluboko do oblasti softwareového inženýrství (a zvyšující tak textový rozsah práce) dává dobrou představu o realizovatelnosti návrhu, myslím, že uvedení informace o částečném ověření reálné implementace popisované metody by nemělo být pro spoluautora generátoru konektorů pro architektury SOFA a CoCoME až tak obtížné.

Odbornou rozpravu při vlastní obhajobě bych rád zaměřil na problematiku efektivního monitoringu a vhodného výběru forem spolupráce jednotlivých částí navrhované architektury.

Doktorand zakládá představu monitoringu na sondách, pracujících se skalárními číselnými indikátory a čítači událostí. Pro hodnocení časového chování sledované aplikace je doplňuje o možnost indikovat jednotlivé události (například jednotlivá volání metod komponent), předávaná informace je pro další zpracování doplněna o časovou známku.

Porovnáme-li navrhované řešení se systémy správy využívanými například v komunikacích (v práci vícekrát zmíněné SNMP), vidíme výrazný rozdíl v doplnění časových známek k záznamům o událostech v aplikaci (u SNMP událostem odpovídají triggerové indikující překročení mezí sledovaných hodnot, u alarmů není přesný časový údaj až tak podstatný a může být získán jako výsledek reakce monitoru). Naproti tomu tabulky s asociativním výběrem, které spolu se stromovou hierarchií spravovaných dat

(MIB) slouží k modelování struktur v SNMP, nejsou pro jednoduchý výkonový monitoring aplikací principiálně nutné (SNMP je používá pro statická a dynamická konfigurační data, např. směrovací tabulky). Když už jsme u hierarchie spravovaných dat, hierarchie jmenných prostorů navrhovaného modelu se od stromové struktury MIB nijak výrazně neliší (až na efektivní vnitřní reprezentaci jednoznačných objektových identifikátorů - OID), výhodou navrhovaného řešení je jistě přímá vazba na strukturu indikátorů implementovaných v nosném operačním systému.

Modernější systémy správy a monitoringu (ale i historický CMS - Common Management System a původní představa OMG) vedle využití jednoduchých skalárních indikátorů a čítačů událostí zahrnují představu objektů jako struktur, se kterými pracují samotné senzory. Výhodou takového přístupu je možnost efektivně implementovat statistiky (u SNMP toho, i když pouze na základě tabulek, využívají statistiky RMON). Typické agregační funkce integrované do sensorů by mohly výrazně zmenšit objem dat v paměti monitorovacího systému.

Vzhledem k představě, že ve fázi vývoje aplikace bude pro programátora klíčové sledování efektivity na úrovni volání metod komponent, stála by za uvážení i možnost integrace alespoň primitivních sensorů a agregačních funkcí do konektorů, vedle zdrojů událostí. Mám dojem, že model, u kterého je, například, monitorovací funkce konektoru schopna sama lokálně filtrovat volání, která jsou zdrojem indikovaných událostí, nebo vytvářet statistiku podle hodnot parametrů a dob odezvy volaných metod, by mohl být v praxi velmi užitečný.

Za analýzu by určitě také stál způsob práce s časem. Návrh předpokládá využití objektu zapouzdřujícího hodinový generátor. Jde o řešení, které pro výkonovou analýzu plně postačuje, architektura poměrně komplikovaných časovačů, jaké používá např. TAU, je pro potřebu pasivního sledování aplikace zbytečná a využije ji pouze strana monitoru. Na druhou stranu, jedná se o extrémně využívanou funkci (záznamy o volání metod budou zřejmě jsou typicky doprovázeny časovými známkami), navíc vyžadující synchronní řešení a proto i s ohledem na její velice jednoduchou implementaci (např. čtení čítače cyklu procesoru několika málo instrukcemi) by stálo za úvahu třeba i její platformově závislé řešení v kódu konektorů.

Návrh monitorovacího systému na řadě míst mluví o aplikovatelnosti modelu na distribuované aplikace, tedy aplikace, u kterých jsou jednotlivé komponenty rozmístěné na různé procesory propojené sítí, a u kterých musíme připustit nezanedbatelné doby komunikace. V této souvislosti by bylo vhodné si uvědomit, že synchronní vazba událostí na čtení hodnot sensorů by mohla vést k výraznému ovlivnění výkonových parametrů aplikace. I přes nutnost zvládnout možnou zvýšenou nekonzistenci informací o právě probíhajícím výpočtu by stálo za úvahu doplnit model o asynchronní získávání dat ze sensorů. Zajímavé by mohlo být i využití reálného paralelismu v moderních procesorech.

Při nasazení monitorovacího systému v distribuovaném prostředí by bylo vhodné alespoň zmínit i problematiku synchronizace času mezi vzdálenými počítači, přestože současné a snadno dostupné technologie, například NTP nebo GPS (každá z nich jiným způsobem), dovolí problém elegantně obejít.

Jak už jsem zmínil, je škoda, že text práce neuvádí alespoň pilotní ověření monitorovacího systému, například formou následné ruční modifikace generovaného kódu konektorů. Za zajímavé bych považoval i ověření vlivu monitoringu na výsledné chování aplikace, i když je jasné, že by si vyžádalo umělé rozšíření počtu sensorů v konektoru.

Ing. Lubomír Bulej má zkušenost s aplikací konektorů v komponentních systémech a této zkušenosti při popisu monitorovací architektury dobře využívá. Své předchozí práce spojené s efektivním generováním konektorů a s problematikou monitorování aplikací na více místech úspěšně publikoval, publikací prošla i v doktorské dizertaci popisovaná metoda implementace monitoringu v konektorech.

Celkově, předkládaná doktorská dizertační práce dobře prokazuje schopnost samostatné vědecké práce Ing. Lubomíra Buleje, řeší aktuální problém s výrazným praktickým dopadem prostředky současné analýzy, a proto ji ve smyslu zákona 172/1990 Sb. doporučuji přijmout k obhajobě s cílem získání titulu PhD. v oboru počítačových věd na MFF UK v Praze.

V Praze dne 22. srpna 2007

