

DINAMIČKI MEHANIZMI TARIFIRANJA I ALOKACIJE CLOUD RESURSA

Branka Mikavica, Aleksandra Kostić-Ljubisavljević
Univerzitet u Beogradu - Saobraćajni fakultet
b.mikavica@sf.bg.ac.rs; a.kostic@sf.bg.ac.rs

Rezime: *Intenzivan razvoj cloud computing-a zahteva unapređivanje odgovarajućih mehanizama tarifiranja i alokacije cloud resursa. Imajući u vidu karakteristike cloud resursa i činjenicu da tražnja za pristup tim resursima nije uniformna, neophodno je primenjivati dinamičke mehanizme za alokaciju i tarifiranje resursa. Ciljevi dinamičkih mehanizama su optimizacija iskorišćenja resursa, maksimiziranje prihoda cloud provajdera, kao i minimiziranje cena koju plaćaju korisnici cloud resursa. Pokazalo se da neki modeli aukcija mogu biti vrlo efikasni, s obzirom da se na taj način resursi dodeljuju korisnicima koji ih najviše vrednuju, a cene se određuju u skladu sa uslovima ponude i tražnje za pristup cloud resursima. U ovom radu je dat pregled nekih predloženih modela aukcija za tarifiranje i alokaciju cloud resursa. Takođe, razmotrene su različite strategije kreiranja ponuda korisnika cloud resursa u procesu aukcija.*

Ključne reči: *cloud, aukcije, tarifiranje, alokacija resursa*

1. Uvod

Cloud computing omogućava pristup skalabilnom i elastičnom skupu deljivih, fizičkih ili virtuelnih resursa. *Cloud* resursi se obezbeđuju u obliku *cloud* instanci, odnosno, virtuelnih mašina koje se odlikuju određenim računarskim i skladišnim kapacitetima. Osnovne prednosti *cloud computing*-a su skalabilnost i niski troškovi. Međutim, postoje i brojni izazovi, kao što su transparentnost pristupa *cloud* resursima, poverenje, pouzdanost, bezbednost, dostupnost, upravljanje resursima i performansama [1]. Presudnu ulogu u odlučivanju za pristup *cloud* resursima ima tarifiranje. Korisnici koji zahtevaju pristup *cloud* resursima plaćaju cene koje zavise od vrste *cloud* instance i primenjenog mehanizma tarifiranja. Takođe, mogući su različiti nivoi garancija u pogledu dostupnosti i terminiranja resursa. U opštem slučaju, tarifni mehanizmi u *cloud* okruženju mogu se klasifikovati na statičke i dinamičke. Statički tarifni mehanizmi su dominantni način tarifiranja u *cloud* okruženju, koji definišu fiksne cene po instanci. Za razliku od statičkih, dinamički mehanizmi omogućavaju pristup *cloud* resursima sve dok je ponuda korisnika *cloud* resursa ista ili viša od cene definisane od strane *cloud* provajdera [2]. Na primer, *cloud* provajder Amazon (*Amazon Elastic Compute Cloud (EC2)*) obezbeđuje tri različite opcije za tarifiranje pristupa *cloud* resursima: rezervisane instance, instance na zahtev i *spot* instance [3]. U slučaju primene mehanizma tarifiranja na zahtev, korisnici pristupaju *cloud*

resursima na zahtev i plaćaju fiksni iznos po obračunskom intervalu bez obaveze sklapanja dugoročnih ugovora sa *cloud* provajderom. Tarifni mehanizam rezervacije odlikuje se naplaćivanjem pretplate za rezervaciju, uz uslov sklapanja ugovorne obaveze, najčešće na jednu godinu ili tri godine. S obzirom da mehanizam tarifiranja na zahtev i mehanizam rezervacije nemaju mogućnost izmene cene u zavisnosti od iskorišćenosti resursa, odnosno, odnosa ponude i tražnje, mogu se smatrati statičkim tarifnim mehanizmima.

Prvi dinamički tarifni mehanizam u *cloud* okruženju uveden je od strane *Amazona* u decembru 2009. godine, pod nazivom *spot* tarifiranje. U slučaju primene ovog tarifnog mehanizma, cene za iznajmljivanje resursa - *spot* instanci, dinamički se određuju od strane *cloud* provajdera i zavise od raspoloživosti neiskorišćenih resursa, kao i ponude i tražnje za *spot* instancama. *Spot* tarifiranje zasniva se na aukciji sa uniformnim određivanjem cena [4].

Korisnici *cloud* resursa koji učestvuju u procesu aukcije kreiraju svoje ponude definisane na osnovu sledećih parametara: vrednosti ponude (maksimalne cene koju su spremni da plate po obračunskom intervalu za datu *spot* instancu), tipa instance, trajanja opsluživanja i geografskog područja koje se opslužuje. Ponude je moguće kreirati u bilo kom trenutku, neograničen broj puta. Prikupljene ponude se sortiraju u opadajućem poretku u odnosu na njihove vrednosti, na osnovu čega *cloud* provajder alokira *spot* instance. Proces se nastavlja sve dok se ne alociraju svi raspoloživi resursi, ili dok se ne zadovolje svi zahtevi korisnika. *Spot* cena je jednaka najnižoj ponudi koja obezbeđuje pravo pristupa resursima. Svi korisnici koji ostvaruju pravo pristupa *cloud* resursima u procesu aukcije plaćaju istu cenu, nezavisno od vrednosti ponude. Korišćenje *spot* instanci je moguće sve dok je vrednost ponude viša od trenutno važeće *spot* cene.

Amazon je 2015. godine uveo novu verziju *spot* instanci - *spot block* instance. Osnovna karakteristika *spot block* instanci je mogućnost rada u kontinuitetu, bez terminiranja, u trajanju od 1 čas do 6 časova. Tarifiranje ovih instanci zasniva se na zahtevanom trajanju rada i dostupnosti resursa. Cene za korišćenje *spot block* instanci, *spot block* cene, najčešće su niže za 30% do 45% od cena *cloud* instanci na zahtev. Kreiranjem ponude za učešće u nadmetanju za ostvarivanje pristupa *spot block* instancama u procesu aukcije, korisnik definiše maksimalnu cenu koju je spreman da plati po času korišćenja *spot block* instance i željeno vreme rada bez terminiranja. Ukoliko postoji dovoljno raspoloživih resursa za zahtevano trajanje rada bez terminiranja, *spot block* instanca biće inicirana, imaće mogućnost kontinuiranog rada bez terminiranja, a tarifiranje će se vršiti na osnovu fiksne cene po času. Po isteku definisanog vremena rada *spot block* instance, instanca biva terminirana od strane *cloud* provajdera. Ovaj model pristupa resursima *cloud* provajdera pogodan je za korisnike koji zahtevaju kontinuiran rad ograničenog trajanja, bez terminiranja.

Za primenu koncepta tarifiranja u zavisnosti od obima korišćenja, koje se sreće u *cloud computing*-u, neophodno je implementirati procese i mehanizme za merenje obima korišćenja, agregirati rezultate merenja, vršiti naplatu i održavati računovodstvene podatke. Korisnici plaćaju fiksnu cenu po jedinici vremena za korišćenje *cloud* resursa po instanci. Međutim, imajući u vidu da tražnja za pristup *cloud* resursima varira, fiksni mehanizmi tarifiranja mogu rezultirati lošim iskorišćenjem resursa. *Cloud* provajderi mogu unaprediti iskorišćenje svojih resursa primenom dinamičkih mehanizama tarifiranja kroz obezbeđivanje *spot* instanci. Na taj način, omogućava se pristup neiskorišćenim resursima po nižim cenama kroz mehanizme aukcija. Međutim, ove instance su pogodne samo korisnicima koji imaju veću toleranciju na kašnjenja i moguće otkaze. Analiza

opravdanosti uvođenja dinamičkog tarifiranja prikazana je u [5]. Uvođenje *spot* instanci može stvoriti cenovnu diskriminaciju, s obzirom da korisnici koji pristup *cloud* resursima vrednuju manje mogu ostvariti pristup *spot* instancama. Takođe, obezbeđivanje *spot* instanci predstavlja jeftiniju alternativu onim korisnicima koji nisu osetljivi na kašnjenje, čime se smanjuje prihod *cloud* provajdera u odnosu na situaciju u kojoj bi se primenjivao samo fiksni mehanizam tarifiranja.

U teoriji, prednosti od primene dinamičkih mehanizama tarifiranja imaju i *cloud* provajderi i korisnici. Korisnici mogu ostvariti niže troškove za pristup *cloud* resursima u periodima nižeg saobraćajnog opterećenja. *Cloud* provajderi mogu podsticati rast tražnje u periodu nižeg saobraćajnog opterećenja, na osnovu čega se postiže bolje iskorišćenje resursa i smanjenje troškova. U *cloud* okruženju, postoji jaka konkurencija ne samo između različitih *cloud* provajdera, već i u vremenu opsluživanja korisnika jednog *cloud* provajdera u situaciji kada korisnici procene da će cene za pristup resursima datog *cloud* provajdera u budućnosti biti niže [6]. Očekuje se da će dinamički tarifni mehanizmi imati sve veću primenu u budućnosti.

Ovaj rad je koncipiran na sledeći način. Nakon uvoda, u drugom delu rada prikazan je pregled nekih predloženih mehanizama aukcija za tarifiranje i alokaciju *cloud* resursa. Prednosti primene mehanizama aukcija, kako sa aspekta *cloud* provajdera, tako i sa aspekta korisnika *cloud* resursa, prikazane su u trećem delu rada. Četvrti deo rada predstavlja analizu mogućih strategija za kreiranje ponuda u procesu aukcija. Izazovi u implementaciji mehanizama aukcija za tarifiranje i alokaciju u *cloud* okruženju prikazani su u petom delu rada. Zaključna razmatranja data su na kraju rada.

2. Pregled nekih predloženih aukcijskih mehanizama u *cloud* okruženju

U literaturi se može naći više predloženih algoritama za tarifiranje i alokaciju resursa u *cloud* okruženju zasnovanih na aukcijama [7-9]. Trenutno najzastupljenija istraživanja u oblasti primene aukcijskih mehanizama u *cloud* okruženju usmerena su na statičke zahteve sa jednom vrstom raspoloživih resursa. Usko povezano sa tarifiranjem pristupa resursima jeste i pitanje alokacije *cloud* resursa. Sveobuhvatan pregled aukcijskih mehanizama za alokaciju *cloud* resursa može se naći u [10, 11]. Postoji više klasifikacija aukcijskih mehanizama u *cloud* okruženju [7]. Neki od tih mehanizama su: *one-sided*, *double-sided* i kombinatorne aukcije.

U *one-sided* aukcijama, ponuđači kreiraju svoje ponude bez uvida u vrednosti ponuda drugih ponuđača. Ponuđač sa najvišom vrednosti ponude ostvaruje mogućnost pristupa resursima. Najpoznatiji aukcijski mehanizam ove vrste je *Vickrey* aukcija, takođe poznata i kao *second-price* aukcija, kada se resursi dodeljuju ponuđaču sa drugom najvišom vrednošću ponude. Vrlo čestu primenu ima uniformna aukcija [4]. U takvoj aukciji, *cloud* provajder alokira resurse korisnicima u opadajućem poretku njihovih ponuda, a korisnici plaćaju iznos koji je jednak vrednosti najniže ponude koja obezbeđuje pravo pristupa resursima u aukciji. Koncept marginalne ponude, odnosno, najviše ponude koja ne obezbeđuje pristup resursima u procesu aukcije za tarifiranje i alokaciju *cloud* resursa pri *second-price* aukciji predložen je u [12]. Taj koncept obezbeđuje visoke prihode *cloud* provajderima, ali istovremeno i malu dobit za korisnike.

U slučaju primene *double-sided* aukcija, i *cloud* provajder i korisnik *cloud* resursa kreiraju svoje ponude. Ukoliko se ponude *cloud* provajdera i korisnika *cloud* resursa mogu kreirati u bilo kom trenutku, radi se o kontinualnim dvostrukim aukcijama. Detaljna analiza

primene *double-sided* mehanizama aukcija u *cloud* okruženju data je u [8]. U *double-sided* aukcijama, korisnici *cloud* resursa mogu zahtevati pristup resursima različitih *cloud* provajdera, dok *cloud* provajderi mogu obezbeđivati pristup resursima brojnim korisnicima. Različiti *cloud* provajderi nude različite tarifne mehanizme. Iz tog razloga, kreiranje adekvatnog zahteva za pristup *cloud* resursima često predstavlja veliki problem korisnicima. Ipak, *double-sided* aukcije mogu biti efikasno rešenje za tarifiranje i alokaciju *cloud* resursa na dvostranim tržištima.

Kombinatorne aukcije često se koriste za tarifiranje i alokaciju *cloud* resursa. U tom slučaju, *cloud* provajderi nude grupe resursa za koje korisnici kreiraju ponude, pri čemu postoji najmanje dva tipa resursa. Kombinatorne aukcije smatraju se pogodnim rešenjem kako za *cloud* provajdere, tako i za korisnike. Problem alokacije nekoliko tipova VM instanci većem broju korisnika koji zahtevaju pristup skupu VM instanci različitih tipova, analiziran je u [13]. Predložena su dva mehanizma alokacije *cloud* resursa zasnovana na kombinatornim aukcijama. Prvi mehanizam razmatra alokaciju nekoliko tipova instanci, pri čemu postoji više raspoloživih instanci po svakom tipu instanci. Drugi mehanizam rešava alokaciju koja se zasniva na vrednostima ponuda korisnika i broju zahtevanih instanci, uzimajući u obzir relativni kapacitet VM instanci. Ova dva mehanizma, dalje su poređena sa mehanizmima alokacije koji se zasnivaju na fiksnim cenama. Pokazalo se da mehanizmi kombinatornih aukcija daju bolja rešenja u odnosu na mehanizme alokacije zasnovane na fiksnim cenama, i to u pogledu iskorišćenja resursa, prihoda i efikasnosti alokacije [13]. Primenom ovih mehanizama moguće je ostvariti više prihode *cloud* provajdera i efikasnije iskorišćenje resursa, uz niže troškove. Međutim, dobit korisnika nije obuhvaćena analizom.

3. Prednosti aukcijskih mehanizama za tarifiranje i alokaciju *cloud* resursa

Uprkos svojoj jednostavnosti, statički mehanizmi tarifiranja ne pružaju podršku maksimizaciji prihoda, s obzirom da neiskorišćeni resursi ne generišu prihode već troškove. Imajući u vidu da tražnja za pristup resursima *cloud* provajdera varira tokom vremena, poželjna je primena dinamičkih tarifnih mehanizama za optimizaciju prihoda *cloud* provajdera. Podešavanje cena na osnovu stvarnog ili prognoziranog odnosa ponude i tražnje može omogućiti izmeštanje tražnje iz perioda visokog opterećenja ka periodu niskog opterećenja. Cilj dinamičkih mehanizama tarifiranja je iznajmljivanje neiskorišćenih resursa po nižim cenama, kako bi se unapredilo iskorišćenje resursa, kao i povećanje cena u slučaju da tražnja za pristupom *cloud* resursima raste. Definisanjem cena u skladu sa realnim ili prognoziranim stanjem odnosa ponude i tražnje, korisnicima se mogu ponuditi neiskorišćeni resursi *cloud* provajdera. Pored toga, moguće je postići i izmeštanje perioda vršnog opterećenja. Kao rezultat, povećava se profit *cloud* provajdera, ali i dobit korisnika. Dinamički oblici tarifiranja mogu se realizovati kroz prikupljanje ponuda od korisnika, na osnovu čega se može vršiti alokacija resursa, a zatim i tarifiranje. Ovakvi tarifni mehanizmi sve više se koriste za alokaciju i tarifiranje neiskorišćenih resursa *cloud* provajdera. Adekvatno postavljene mehanizmi aukcija mogu biti vrlo efikasno rešenje, s obzirom na to da se resursi alociraju korisnicima koji ih najviše vrednuju. Na taj način, stvaraju se uslovi za razvoj konkurencije između korisnika, a korisnici se podstiču da kreiraju ponude čije su vrednosti u skladu sa realnim vrednostima zahtevanih resursa [14].

Cloud provajderi nude skup različitih računarskih i skladišnih resursa različitim grupama korisnika primenom različitih tarifnih mehanizama. U skladu s tim, iz perspektive korisnika *cloud* resursa, adekvatan tarifni mehanizam podrazumeva efikasno obezbeđivanje resursa po prihvatljivoj ceni. Iz perspektive *cloud* provajdera, efikasan tarifni mehanizam obezbeđuje maksimizaciju profita uz optimizaciju iskorišćenosti resursa. Zbog svojih prednosti, različiti mehanizmi aukcija mogu biti pogodni za realizaciju dinamičkog tarifiranja u *cloud* okruženju [14]. Aukcije podstiču korisnike *cloud* resursa da biraju vrednosti ponuda koje su zaista spremni da plate za zahtevane resurse, obezbeđuju da se resursi alociraju korisnicima koji ih najviše vrednuju [13], a cene se koriguju u skladu sa odnosom ponude i tražnje na osnovu podsticanja konkurencije između korisnika. Efikasnost iskorišćenja *cloud* resursa je unapređena primenom aukcijskih mehanizama za tarifiranje i alokaciju. Stoga, takvi tarifni mehanizmi mogu efikasno odrediti tržišnu vrednost *cloud* resursa, naročito u slučaju alokacije relativno ograničenih resursa potencijalno velikom broju korisnika.

Posmatrano iz ugla korisnika *cloud* resursa, aukcijski mehanizmi tarifiranja definišu, u proseku, niže cene u poređenju sa fiksnim mehanizmima tarifiranja. *Spot* tarifiranje, kao tipičan primer tarifiranja zasnovanog na aukcijama, obezbeđuje značajne uštede u slučaju kada se koriste *spot* instance umesto instanci na zahtev. Neke empirijske studije pokazuju da uz adekvatno definisane ponude korisnika, uštede na osnovu korišćenja *spot* instanci umesto instanci na zahtev mogu biti i do 36%. Pored toga, *spot* tarifiranje unapređuje iskustveni kvalitet. U situacijama kada raste tražnja za pristupom rezervisanim i instancama na zahtev, korisnici mogu smanjiti vreme odziva korišćenjem alternativnih instanci, *spot* instanci ili njihove varijacije - *spot block* instanci. Takođe, kombinatorne aukcije obezbeđuju pogodnije uslove za alokaciju različitih tipova *cloud* resursa korisnicima sa visokim zahtevima. Uz odgovarajuću tržišnu strukturu, moguće je ostvariti pravične relacije između *cloud* provajdera i korisnika kroz mehanizme aukcija. Na taj način se ostvaruje veća društvena dobit na nivou celog *cloud* okruženja [15].

4. Strategije kreiranja ponuda *cloud* korisnika u procesu aukcija

Izbor odgovarajuće strategije pri kreiranju ponuda u aukcijskim mehanizmima tarifiranja i alokacije resursa u *cloud* okruženju, od ključnog je značaja iz perspektive korisnika *cloud* resursa. *Cloud* provajder alokira neiskorišćene resurse korisnicima u zavisnosti od svih pristiglih zahteva sa ponudama i raspoloživosti resursa. Nakon predaje korisničkih ponuda, *cloud* provajder agregira ponude i tražnju za pristupom *cloud* resursima i upoređuje je sa ponudom. Angažovanje neiskorišćenih resursa *cloud* provajdera ostvaruje značajne uštede u troškovima kako za *cloud* provajdere, tako i za korisnike. Međutim, izbor odgovarajuće strategije pri kreiranju ponuda je od presudnog značaja za korisnike. U opštem slučaju, izabrana strategija utiče na troškove, na vreme čekanja na opsluživanje i stopu terminiranja. Ponude niže vrednosti smanjuju troškove. Međutim, druge mere performansi, kao što su vreme opsluge i stopa terminiranja, bivaju narušene. Iz tog razloga, *cloud* provajderi često preporučuju izbor strategije kojom se kreiraju visoke vrednosti ponuda, kako bi se smanjila verovatnoća terminiranja *cloud* instance pre završetka opsluživanja. Korisnici koji su spremni da plate samo niže cene za pristup *cloud* resursima najčešće kreiraju ponude niskih vrednosti. Međutim, *cloud* instance koje su inicirane od strane korisnika sa niskom vrednošću ponude imaju manju raspoloživost i veliku verovatnoću terminiranja u slučaju kada raste tržišna cena pristupa *cloud* resursima.

Česta terminiranja mogu dovesti do otkaza, produženog vremena opsluge i sveukupnog narušavanja performansi. Dakle, strategije kreiranja ponuda predstavljaju kompromis između troškova i raspoloživosti, odnosno, stope terminiranja. Adekvatna strategija kreiranja ponuda podrazumeva strategiju koja minimizira troškove uz zadovoljavanje različitih ograničenja, kao što je zahtevano vreme opsluživanja. Složenost adekvatnog izbora strategije je razlog što, uprkos nižim cenama, pristup *cloud* resursima primenom aukcija još uvek nema rasprostranjenu primenu [16].

Nekoliko strategija kreiranja ponuda analizirano je u [16]: kreiranje ponuda čija je vrednost približna vrednosti cene rezervisanih instanci, kreiranje ponuda čija je vrednost viša od vrednosti cene *spot* instanci, kreiranje ponuda čija je vrednost približna vrednosti cene instanci na zahtev i kreiranje ponuda čija je vrednost viša od vrednosti cene instanci na zahtev. Ta studija je pokazala da se izborom strategije kreiranja ponuda čija je vrednost približna vrednosti cene rezervisanih instanci smanjuju troškovi uz duže čekanje na opslugu i višu stopu terminiranja instance. Takođe, usvajanjem strategije kreiranja ponuda čija je vrednost je viša od cene instance na zahtev ostvaruje se najniže vreme čekanja na opsluživanje i najniža stopa terminiranja, ali i viši troškovi u odnosu na druge strategije kreiranja ponuda. Pokazalo se i da je rast marginalnih troškova značajno viši od smanjenja vremena čekanja na opslugu i smanjenja stope terminiranja. Međutim, 35% korisnika bira strategiju kreiranja ponuda čija je vrednost je iznad cene instance na zahtev. Kreiranje ponuda čija je vrednost viša od prosečne *spot* cene, kao i kreiranje ponuda čija je vrednost približno jednaka ceni instance na zahtev obezbeđuje zadovoljavajući kompromis između troškova, vremena čekanja na opsluživanje i terminiranja. Dakle, strategija kreiranja ponuda čija je vrednost je približno jednaka ceni rezervisanih instanci pogodna je za korisnike *cloud* resursa koji su spremni da plate niske cene za rezervaciju resursa i imaju fleksibilne rokove za opsluživanje. Strategije kreiranja ponuda sa vrednostima koje su više od srednje vrednosti *spot* cene i strategije kreiranja ponuda sa vrednostima koje su približno jednake ceni instance na zahtev su prihvatljivije za korisnike koji imaju iste zahteve u pogledu troškova, vremena čekanja na opsluživanje i terminiranja.

Poređenje ponuda u mehanizmu aukcija u slučaju alokacije *spot block* instanci složenije je u odnosu na poređenje ponuda u slučaju alokacije regularnih *spot* instanci. Složenost je povećana usled karakteristika *spot block* instanci da rade bez terminiranja čak i kada cene postanu više od vrednosti ponuda. S obzirom na to da se cene *spot block* instanci razlikuju u zavisnosti od zahtevanog vremena rada bez terminiranja, ponude se ne mogu jednostavno sortirati u zavisnosti od njihove vrednosti [17]. U [18], predložen je model za translaciju ponuda u jedinične ponude, a sve u cilju omogućavanja njihovog poređenja u procesu aukcije.

Optimizacija odnosa troškovi - raspoloživost *cloud* resursa može se ostvariti kroz različite strategije kreiranja ponuda. Kada korisnik povećava vrednost svoje ponude, raspoloživost date instance se povećava na račun povećane cene po satu korišćenja. Takođe, kreiranje ponuda sa velikim rasponom vrednosti ne utiče na raspoloživost neiskorišćenih *cloud* resursa. Isto važi i za troškove, s obzirom da u aukciji ne postoje penali za kreiranje ponuda previsokih vrednosti. Još jedan važan parametar je učestanost terminiranja. Nezavisno od vrednosti ponude, terminiranja su neizbežna pri pristupanju neiskorišćenim *cloud* resursima. U skladu s tim, neke studije su pokazale da korisnici ne moraju birati neke složene strategije pri kreiranju ponuda [16]. Umesto toga, korisnici treba da biraju odgovarajući tip instance kako bi se smanjila verovatnoća terminiranja, pri čemu vrednost ponude treba da bude približno jednaka ceni instance na zahtev. Ipak, poželjno je

birati različite tipove instanci kako bi bilo moguće izvršiti migraciju između instanci po potrebi. Na taj način, moguće je značajno smanjiti verovatnoću otkaza usled terminiranja i istovremeno smanjiti troškove. Efikasnost iskorišćenja *cloud* resursa takođe može biti unapređena. Dakle, odgovarajuća kombinacija tipova instanci i primenjena politika tolerancije na otkaze su ključni faktori koji određuju performanse i troškove u *cloud* okruženju.

5. Izazovi u primeni mehanizama aukcija za tarifiranje i alokaciju *cloud* resursa

Problemi optimizacije iskorišćenja *cloud* resursa i maksimizacije prihoda *cloud* provajdera su NP-kompletni problemi [19]. Implementacija takvih mehanizama za tarifiranje i alokaciju resursa je vrlo zahtevna. Pored toga, neistinito kreiranje ponuda i međusobna saradnja između korisnika *cloud* resursa mogu dovesti do toga da proces tarifiranja primenom aukcija postane neprofitabilan za *cloud* provajdere. Korisnici *cloud* resursa treba da prepoznaju i kontrolišu potencijalne izvore koji mogu uticati na proces donošenja odluka o vrednosti ponuda. Ti izvori mogu biti preporuke *cloud* provajdera, uputstva za korisnike, vrednosti *spot* cena u prethodnom periodu, vrednosti ponuda drugih korisnika itd. Kako bi mehanizam aukcije bio efikasan, korisnici treba da definišu ponude koje približno oslikavaju njihove troškove, toleranciju na vreme čekanja na opslugu i toleranciju na potencijalna terminiranja *cloud* instanci. Neke studije pokazuju da kreiranje ponuda iznad cena instanci na zahtev nije ekonomski opravdano [20].

Jedan od najznačajnijih zahteva koji mehanizam aukcije treba da zadovolji tiče se istinitosti. Mehanizam aukcije je istinit ako učesnici u aukciji ne mogu povećati svoje prihode na osnovu povlačenja poteza (kreiranja ponuda) koji ne predstavljaju njihova istinita vrednovanja. Takav istinit mehanizam aukcija za tarifiranje i alokaciju resursa je moguć samo ako su poznate sve informacije o vrednostima ponuda drugih korisnika.

Implementacija mehanizma koji bi zadovoljio zahteve i *cloud* provajdera i korisnika *cloud* resursa od velikog je značaja. Moguće je uspostaviti istinit mehanizam aukcije za tarifiranje i alokaciju *cloud* resursa, takav da i *cloud* provajderi i korisnici *cloud* resursa imaju koristi. U realnim situacijama, korisnici *cloud* resursa retko znaju koliko resursa im je potrebno za opsluživanje njihovih zahteva. Iz tog razloga, zahtevi korisnika često su nekonzistentni sa resursima provajdera. Efikasno rešenje ovog problema može biti istinita dvostruka aukcija.

Korisnici *cloud* resursa suočavaju se sa izvesnim ograničenjima u procesu aukcija za pristup neiskorišćenim resursima. Najznačajnije ograničenje odnosi se na dostupnost i pouzdanost *cloud* instanci. *Cloud* provajder kontroliše terminiranje neiskorišćenih *cloud* resursa u zavisnosti od odnosa između ponude i tražnje za pristupom *cloud* resursima i u skladu s tim koriguje cene. Na osnovu toga, *cloud* instance se mogu terminirati u bilo kom trenutku. To može imati negativan efekat na performanse, što utiče i na ekonomsku isplativost pristupa neiskorišćenim *cloud* resursima. Takođe, relativno često terminiranje instanci može biti jednako neefikasno kao i statički mehanizmi tarifiranja [15]. Može se zaključiti da su dinamički mehanizmi tarifiranja neadekvatni za opsluživanje zahteva koji se odlikuju malom fleksibilnošću i dugim vremenom opsluge. Još jedno izuzetno važno pitanje u kontekstu aukcijskih mehanizama tarifiranja jeste odgovarajuće kreiranje ponuda. Kreiranje ponuda čija je vrednost niska može rezultovati vrlo dugim vremenima opsluge bez značajnih ušteda u troškovima. Za razliku od toga, kreiranje ponuda čije su vrednosti velike može dovesti do značajnog rasta troškova bez smanjivanja vremena opsluge.

Većina predloženih aukcijskih mehanizama tarifiranja i alokacije resursa razmatra samo jedan atribut *cloud* resursa a korisnici pri kreiranju ponuda definišu samo vrednost cene. Parametri, kao što su reputacija, iskustvo korisnika itd, takođe treba da budu uključeni u analizu. Mehanizam aukcije sa više posmatranih atributa može biti pogodno rešenje koje omogućava korisnicima *cloud* resursa da kreiranju ponude u vezi sa različitim atributima *cloud* resursa zajedno sa cenom. U tom slučaju, *cloud* provajder bira pobjednika u aukciji na osnovu definisanih atributa. *Cloud* provajder može nuditi skup *cloud* resursa u formi paketa. Grupisanje *cloud* resursa tako da se zadovolje zahtevi korisnika i ostvari maksimizacija profita *cloud* provajdera predstavlja veliki izazov. U takvom slučaju, kombinatorne aukcije predstavljaju obećavajuće rešenje, s obzirom na to da korisnici mogu kreirati ponude za grupu zavisnih i komplementarnih resursa. Ako se posmatra više *cloud* provajdera i više korisnika *cloud* resursa, dvostruke aukcije smatraju se pogodnim rešenjem zbog veće efikasnosti u poređenju sa jednostrukim aukcijama. Tarifiranje i alokacija *cloud* resursa izvršava se kada se pojavi zahtev korisnika za pristup *cloud* resursima, ili kada *cloud* resursi bivaju ponovo dostupni. Svaki korisnik ima ograničen budžet u određenom vremenskom intervalu. Iz tih razloga, prednost se daje *online* mehanizmima aukcija, imajući u vidu da ova vrsta aukcija uzima u obzir i vremensku karakteristiku pristupa *cloud* resursima.

6. Zaključak

Tarifiranje i alokacija *cloud* resursa primenom različitih mehanizama aukcija predmet je istraživanja brojnih studija. Dinamički mehanizmi tarifiranja i alokacije resursa obezbeđuju brojne prednosti koje su uticale na to da vodeći *cloud* provajderi ponude pristup svojim resursima kroz mehanizme aukcija. Aukcije se smatraju pogodnim rešenjem za maksimiziranje prihoda *cloud* provajdera. Takođe, podstiču korisnike *cloud* resursa na kreiranje istinitih vrednosti ponuda. Tarifiranje, kreiranje ponuda i alokacija *cloud* resursa zahtevaju sveobuhvatan i multidisciplinarn pristup. Zahteva se modelovanje i prognoziranje cena, analiza strategija kreiranja ponuda, nove metode za unapređenje dostupnosti resursa, smanjenje verovatnoće otkaza itd. U ovom radu, prikazane su najznačajnije prednosti aukcijskih mehanizama za tarifiranje i alokaciju *cloud* resursa. Analiziran je značaj izbora adekvatne strategije za kreiranje ponuda u procesu aukcija. Uprkos brojnim prednostima, primena dinamičkih mehanizama postavlja nove zahteve za istraživanje. Neki od izazova i potencijalna rešenja opisani su u ovom radu.

Zahvalnica

Ovaj rad je deo istraživanja u okviru projekta TR32025 Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije.

Literatura

- [1]. N. Gullhav, B. Nygreen, „A Branch and Price Approach for Deployment of Multi-Tier Software services in Clouds“, *Computers & Operations Research*, 75, pp. 12-27, 2016.

- [2]. F. Alzhouri, A. Agarwal, „Dynamic Pricing Scheme: Towards Cloud Revenue Maximization“, *In Proceedings of the IEEE 7th International Conference on Cloud Computing Technology and Science*, Vancouver, Canada, pp. 168-173, 2015.
- [3]. B. Mikavica, V. Radonjić Đogatović, A. Kostić-Ljubisavljević, „Mogućnosti primene spot pricing mehanizma u cloud okruženju“, *Zbornik radova XXXIV Simpozijuma o novim tehnologijama u poštanskom i telekomunikacionom saobraćaju – POSTEL 2016*, Beograd, str. 305-314, 2016.
- [4]. Q. Zhang, E. Gurses, R. Boutaba, „Dynamic Resource Allocation for Spot Markets in Clouds“, *in Proceedings of 4th IEEE International Conference on Utility and Cloud Computing (UCC)*, Victoria, pp. 178–185, 2011.
- [5]. V. Abhishek, I. A. Kash, P. Key, „Fixed and Market Pricing for Cloud Services“, *in Proceedings of IEEE INFOCOM Workshops*, Orlando, pp. 157–162, 2012.
- [6]. J. Wienman, „Cloud Pricing and Markets“, *IEEE Cloud Computing*, 2(1), pp. 10-13, 2015.
- [7]. G. Baranwal, D. Kumar, Z. Raza, D. P. Vidyarthi, *Auction Based Resource Provisioning in Cloud Computing*, Springer Nature Singapore, Singapore, 2018.
- [8]. Kumar, G. Baranwal, Z. Raza, D. P. Vidyarthi, „A Systematic Study of Double Auction Mechanisms in Cloud Computing“, *Journal of Systems and Software*, 125, pp. 234-255, 2017.
- [9]. L. Lu, J. Yu, Y. Zhu, M. Li, „A Double Auction Mechanisms to Bridge Users' Task Requirements and Providers' Resources in Two-Sided Cloud Markets“, *IEEE Transaction on Parallel and Distributed Systems*, 29(4), pp. 720-730, 2018.
- [10]. F. Sheikholeslami, N. J. Navimipour, „Auction- Based Resource Allocation Mechanisms in the Cloud Environment: A Review of the Literature and Reflection on Future Challenges“, *Concurrency and Computation Practice and Experience*, 30(16), pp. 1-15, 2018.
- [11]. B. Mikavica, A. Kostic-Ljubisavljevic, V. Radonjić Đogatović, „Mogućnosti primene aukcija za tarifiranje i alokaciju resursa cloud provajdera“, *Tehnika*, 6, pp. 879-885, 2017.
- [12]. W. -Y. Lin, G. -Y. Lin, H. -Y. Wei, „Dynamic Auction Mechanism for Cloud Resource Allocation“, *in Proceedings of the 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (CCGrid)*, Melbourne, pp. 591–592, 2010.
- [13]. S. Zaman, D. Grosu, „Combinatorial Auction-Based Allocation of Virtual Instances in Clouds“, *Journal of Parallel and Distributed Computing*, 73(4), pp. 495-508, 2013.
- [14]. N. Toosi, F. Khodadadi, R. Buyya, „An Auction Mechanism for Cloud Spot Markets“, *ACM Transactions on Autonomous and Adaptive Systems*, 11(1), pp. 25–57, 2016.
- [15]. W. Shi, L. Zhang, C. Wu, Z. Li, F. C. M. Lau, „An Online Auction Framework for Dynamic Resource Provisioning in Cloud Computing“, *IEEE/ACM Transactions on Networking*, 24(4), pp. 2060-2073, 2016.
- [16]. P. Sharma, D. Irwin, P. Shenov, „Keep It Simple: Bidding for Servers in Today's Cloud Platforms“, *IEEE Internet Computing*, 21(3), pp. 88-92, 2017.
- [17]. B. Mikavica, *Modeli vertikalne interkonekcije u mrežama budućeg Interneta*, doktorska disertacija, Beograd, 2019.

- [18]. B. Mikavica, A. Kostic-Ljubisavljevic, „Pricing and Bidding Strategies for Cloud Spot Block Instances“, in *Proceedings of the 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, Hrvatska, pp. 419-424, 2018.
- [19]. K. Song, Y. Yao, L. Golubchik, „Exploring the Profit-Reliability Trade-off in Amazon's Spot Instance Market: A Better Pricing Mechanism“, in *Proceedings of the 21st IEEE/ACM International Symposium on Quality of Service (IWQoS 2013)*, Montreal, Canada, pp. 1-10, 2013.
- [20]. S. Karunakaran, R. Sundarraj, „Bidding Strategies for Spot Instances in Cloud Computing Markets“, *IEEE Internet Computing*, 19(3), pp. 32-40, 2014.

Abstract: *The permanent growth in cloud computing deployment requires improvements in pricing and resource allocation mechanisms. Considering the characteristics of cloud resources and the fact that demand is not uniform, dynamic resource allocation and pricing mechanism are necessary. Goals of dynamic mechanisms are the optimization of resource utilization, cloud providers' revenues maximization, as well as the minimization of prices for cloud customers. It is shown that some auction-based mechanisms can be very efficient. The resources are assigned to the customers that value the access the most, while prices are set according to the demand and supply relations. This paper presents an overview of some proposed auction-based models for cloud resources' pricing and allocation. Also, various cloud customers' bidding strategies in the auction process are addressed.*

Keywords: *cloud, auctions, pricing, resource allocation*

**DYNAMIC MECHANISMS FOR PRICING
AND ALLOCATION OF CLOUD RESOURCES**
Branka Mikavica, Aleksandra Kostić-Ljubisavljević