

# تاثیر عرضه پله‌ای تولید در استراتژی قیمت‌دهی واحد تولیدی در بازار انرژی و رزرو بر پایه یادگیری تقویتی

سید محمد علی ناصری جوارشک، سمیه حسن پور دربان، امین نوری

دانشکده برق و مهندسی پزشکی

دانشگاه صنعتی سجاد

مشهد، ایران

[S.hasanpour@sadjad.ac.ir](mailto:S.hasanpour@sadjad.ac.ir), [Amin.noori@sadjad.ac.ir](mailto:Amin.noori@sadjad.ac.ir), [A.naseri1369@sadjad.ac.ir](mailto:A.naseri1369@sadjad.ac.ir)

## ۱. مقدمه

در دو دهه گذشته صنعت برق بسیاری از کشورها تحت فرآیند تجدید ساختار در حال حرکت به سمت تجاری‌سازی و رقابت قرار گرفته است. تحت این محیط رقابتی تمام تولیدکنندگان در تلاش برای کسب سود بیشتر و سهم بالاتر از بازار هستند.

سیستم‌های قدرت همواره نیازمند قابلیت اطمینان کافی برای جلوگیری از اتفاقات پیش‌بینی نشده مانند قطع تولید و تغییرات ناگهانی بار می‌باشند. بازار خدمات جانبی بخشی جدایی ناپذیر در بازار برق محسوب می‌شوند که علاوه بر تامین بخشی از قابلیت اطمینان سیستم، سبب افزایش سودمندی و بهره‌وری شرکت‌کنندگان در بازار خواهد شد [1].

در اکثر سیستم‌های تجدید ساختار یافته مانند کالیفرنیا، نیویورک، New England، تامین خدمات جانبی بر عهده بهره‌بردار مستقل سیستم (ISO)<sup>1</sup> است. یکی از این خدمات جانبی، ظرفیت رزرو چرخشی می‌باشد که مسئله کمبود توان غیر قابل پیش‌بینی که تهدیدی برای پایداری سیستم به‌شمار می‌آید را حل می‌کند [2]. از دیدگاه بازار، تخصیص انرژی و رزرو چرخشی به دو صورت پشت سرهم [3] و یا همزمان [4] صورت می‌گیرد. در تخصیص پشت سرهم، بازار انرژی در اولویت قرار دارد، سپس بازار رزرو

چکیده — با ظهور پدیده تجدید ساختار در صنعت برق، قوانین پیچیده‌ای برای طراحی بازار انرژی و انواع سرویس‌های جانبی توسط تصمیم‌گیران بازار ارائه شده است. انتخاب روش مناسب عرضه تولید یکی از موارد کلیدی در طراحی بازار است که تاثیر زیادی در کسب سود بیشتر برای شرکت‌کنندگان در بازار دارد. عرضه پله‌ای تولید یکی از روش‌های جدید برای این منظور، در مناقصات بازار است. در این مقاله، تامین‌کننده‌های انرژی به عنوان عامل یادگیرنده در مساله الگوریتم یادگیری تقویتی در نظر گرفته شده‌اند و با تعاملی که با محیط و مصرف‌کننده‌ها دارند یاد می‌گیرند تا قیمت را به گونه‌ای پیشنهاد دهند که بیشترین سود ممکن را در دراز مدت بدست آورند. یادگیری از طریق الگوریتم یادگیری-Q در دو حالت عرضه تک پله‌ای و سه پله‌ای تولید در بازار انرژی و رزرو انجام شده است و در نهایت نتایج شبیه‌سازی تاثیر عرضه سه پله‌ای بر پایه یادگیری-Q در ماکزیمم کردن سود حاصل از بازار انرژی و رزرو را به خوبی نشان می‌دهد.

واژه‌های کلیدی — تجدید ساختار؛ سرویس‌های جانبی؛ بازار انرژی و رزرو؛ یادگیری-Q؛

<sup>1</sup> Independent system operator

این امر از الگوریتم یادگیری-Q برای یافتن استراتژی قیمت‌دهی واحدهای تولیدی استفاده شده است.

## ۲. ساختار بازار انرژی و رزرو

در این مقاله بازار مورد نظر از نوع حوضچه توان<sup>۶</sup> در نظر گرفته شده است و در یک دوره ۲۴ ساعته و تحت قانون قیمت گذاری PAB اجرا می‌شود. در قیمت گذاری PAB به برندگان بازار بر اساس قیمت پیشنهادی خود پرداخت سود انجام می‌گیرد. ISO مسئله بهینه‌سازی برای حداکثر کردن رفاه اجتماعی حل می‌کند [5]:

$$F = \max \left[ \sum_{j=1}^D (C_{e,j} \cdot P_j + C_{R,j} \cdot R_j) - \sum_{i=1}^G (C_{e,i} \cdot P_i + C_{R,i} \cdot R_i) \right] \quad (1)$$

که در آن  $C_{e,i}$  و  $C_{e,j}$  قیمت پیشنهادی خرید و فروش برای انرژی  $\$/MWh$ ،  $C_{R,i}$  و  $C_{R,j}$  قیمت پیشنهادی خرید و فروش برای رزرو  $\$/MWh$ ،  $P_i$  و  $P_j$  مقدار انرژی مصرفی و تولیدی  $MWh$ ،  $R_i$  و  $R_j$  مقدار رزرو مصرفی و تولیدی  $MWh$ ،  $D$  و  $G$  تعداد خریداران و فروشندگان هستند.

از طرفی مصرف‌کنندگان عملاً از روند قیمت گذاری مورد مطالعه حذف می‌شود. در نتیجه معیار حداکثر رفاه اجتماعی به دلیل عدم وجود یکی از طرفین تغییر می‌کند. بنابراین باید (۱) به معیار کمترین هزینه انرژی و رزرو اصلاح شود [5]:

$$F = \min \left[ \sum_{i=1}^D (C_{e,i} \cdot P_i + C_{R,i} \cdot R_i \cdot I(C_R, MCP_R)) \right] \quad (2)$$

$I(C_R, MCP_R)$  تابع احتمالی برای بیان حضور یا عدم حضور ظرفیت رزرو در بازار استفاده می‌شود و بصورت (۳) قابل محاسبه است.

$$I(C_R, MCP_R) = \begin{cases} 1 & C_R < MCP_R \\ 0 & C_R > MCP_R \end{cases} \quad (3)$$

$MCP_R$  قیمت نسویه بازار رزرو است که توسط ISO مشخص می‌شود.

مدل ارائه شده برای بازار رزرو و انرژی در این مقاله (A|L) خواهد بود

جهت تامین قابلیت اطمینان سیستم اجرا می‌شود. در تخصیص همزمان، تسویه بازار انرژی و رزرو بصورت همزمان صورت می‌گیرد،

حل مسئله بهینه سازی تخصیص انرژی و رزرو، جزو توابع عملیاتی بازار محسوب می‌شوند که توسط ISO اجرا می‌شود. در مرجع [3] فرمولبندی مسئله بهینه‌سازی بصورت یک مسئله احتمالاتی به منظور بررسی حضور خرده فروشان (Disco)<sup>۲</sup> در بازار رزرو به عنوان یک بار قابل قطع جهت تامین رزرو بیان شده است. در مرجع [5] مسئله بهینه‌سازی با هدف حداقل کردن هزینه‌های تولید مطرح شده است. حداکثر کردن سود مسئله بهینه‌سازی تخصیص انرژی و رزرو در مرجع [6] بیان شده است.

در محیط تجدید ساختار یافته، استراتژی قیمت‌دهی به یکی از مهمترین مسائل برای تولیدکنندگان تبدیل شده است. روش‌های متنوعی برای قیمت‌دهی در مقالات مختلف ارائه شده است. یکی از محبوبترین این روش‌ها الگوریتم یادگیری-Q است. این الگوریتم بر پایه تعاملی که با محیط انجام می‌دهد سیاست بهینه را پیدا می‌کند، که در این مساله سود بیشتر را برای تولید کننده به همراه دارد.

در مقالات مختلف از الگوریتم یادگیری-Q به شکل‌های گوناگون استفاده شده است. در مرجع [7] از الگوریتم یادگیری-Q برای بررسی تأثیر رفتار استراتژیک شرکت‌کنندگان بازار روی قیمت بازار استفاده کرده است، در حالی که فرض شده هیچ‌کدام از بازیگران قدرت بازار را ندارند. در مراجع [8,9] از این الگوریتم برای مقایسه دو نوع قانون قیمت‌گذاری یکنواخت<sup>۳</sup> و پرداخت براساس پیشنهاد (PAB)<sup>۴</sup> استفاده کرده‌اند. الگوریتم SA-Q-Learning<sup>۵</sup> برای پیدا کردن استراتژی بهینه در طول فرآیند یادگیری در [10] بیان شده است.

در ساختار برق رقابتی با توجه به متغیر بودن هزینه‌ها و درآمدها، تولیدکنندگان برای کسب سود بیشتر و هوشمندانه‌تر از روش پله‌ای برای عرضه تولید خود در بازار استفاده می‌کنند. در قیمت‌دهی پله‌ای تولیدکننده ظرفیت تولید خود را در مقادیر و قیمت‌های مختلف پیشنهاد می‌دهد. در نتیجه تولیدکنندگان درمی‌یابند که اگر قیمت بالاتری در بازار پیشنهاد دهند، شانس کسب سود بیشتری را دارند.

در این مقاله قصد برآن است تا تأثیر عرضه پله‌ای تولید را بر رفتار تولیدکنندگان بازار انرژی و رزرو مورد تجزیه و تحلیل قرارگیرد. برای تحقق

<sup>2</sup> Distribution Company

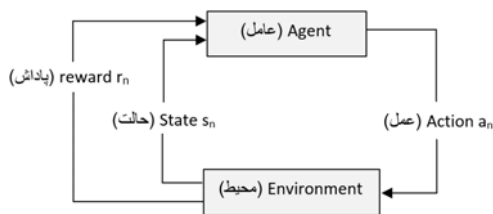
<sup>3</sup> Uniform

<sup>4</sup> Pay as bid

<sup>5</sup> Simulation annealing Q-Learning

<sup>6</sup> Pool

<sup>7</sup> Market clearing price



شکل (۱): روند یادگیری عامل از طریق تعامل با محیط

الگوریتم یادگیری-Q نوعی الگوریتم یادگیری تقویتی است، که توسط واتکینز در سال ۱۹۹۸ برای حل مسائل مارکوف با اطلاعات ناقص ارائه شده است. نقاط قوت این الگوریتم عدم نیاز به مدل دقیق از محیط و استفادهی برخط برای پیدا کردن استراتژی بهینه است. این الگوریتم روش خوبی برای حل مشکلات تصمیم‌گیری‌های پی‌درپی می‌باشد، که در آن عمل انجام شده به دنبال ای از تصمیم‌گیری‌ها ساخته می‌شود [9].

همانطور که در شکل (۱) مشاهده می‌شود، برای بیان فرمول‌بندی الگوریتم یادگیری-Q اگر  $S = \{s_1, s_2, s_3, \dots, s_n\}$  و  $A = \{a_1, a_2, a_3, \dots, a_n\}$  مجموعه‌ای محدود از حالت‌های محیط و عمل‌هایی باشد که عامل انجام می‌دهد، در هر مرحله زمانی  $t$ ، هر عامل بر اساس حالت  $s_t$  که در آن قرار دارد عمل  $a_t$  را انتخاب می‌کند. در این حالت عامل پاداش فوری  $r_{t+1}$  را دریافت می‌کند و به حالت جدید می‌رود. در نهایت تابع تقریبی ارزش با استفاده از یک جدول حالت-عمل مشخص می‌شود، که مقادیر آن در جدولی تحت عنوان جدول Q تعریف می‌شوند [8]. هر عامل برای پیدا کردن بهترین سیاست با هدف حداکثر کردن مقدار Q هر حالت در هر دوره زمانی یا حداکثر کردن پاداش کل تلاش می‌کند. در هر مرحله مقدار Q از رابطه (۶) بروزرسانی می‌شود [7-9].

$$Q(s_t, a_t) = (1 - \alpha_t) Q(s_t, a_t) + \alpha_t [r_{t+1} + \gamma \max_a Q(s_{t+1}, a)] \quad (6)$$

$\alpha_t$  نرخ یادگیری، میزان یادگیری عامل از اطلاعات جدید بدست آمده از مرحله قبل است که مقدار آن بین ۰ و ۱ است.  $\alpha_t = 0$  یعنی جدول Q هرگز بروز رسانی نمی‌شود و عامل یادگیری ندارد.  $\alpha_t = 1$  یعنی یادگیری به آخرین پاداش آنی که دریافت کرده است و ارزش‌های حالت-عمل‌های بعدی اهمیت می‌دهد. که برای یادگیری مناسب باید تعاملی بین یادگیری‌هایی که تا الان انجام شده است و آخرین داده یادگیری برقرار کنیم. در مرحله یادگیری نرخ یادگیری  $\alpha(s, a)$  مطابق (۷) با تعداد تکرار که یک جفت حالت-عمل تکرار شده، نسبت عکس دارد. در مرحله اصلی قیمت دهی مقدار  $\alpha_t$  با توجه به پارامترهای تنظیمی عامل‌ها مشخص می‌شود.  $\gamma$  نرخ

که در آن تولیدکننده تنها هزینه آمادگی یا هزینه از دست رفته را دریافت می‌کند [11]. هزینه آمادگی یعنی مبلغی که تولیدکننده به ازای ظرفیت رزرو تخصیص یافته به آن دریافت می‌کند. اما اگر تولیدکننده نتواند مقدار ظرفیت رزرو خود را به فروش برساند، به دلیل محروم ماندن از بازار انرژی و کاهش خسارت ناشی از آن، هزینه فرصت از دست رفته را دریافت خواهد کرد. هزینه فرصت از دست رفته بصورت (۴) بیان می‌شود [11, 12].

$$f(C_{R,i}, MCP_R) = \begin{cases} \max[0, (MCP_R - C_{R,i})(\bar{P} - P_i)] & C_{R,i} < MCP_R \\ 0 & C_{R,i} \geq MCP_R \end{cases} \quad (4)$$

$\bar{P}$  نتیجه بهینه سازی بدون در نظر گرفتن قیود تامین رزرو می‌باشد.

در نهایت این مسئله تحت مدل (A|L) برای هر ساعت بهره‌برداری بصورت (۵) قابل بیان است. (۵-۱) تابع هدف و (۵-۲) تا (۵-۶) قیود مسئله بهینه‌سازی را نشان می‌دهند.

$$F = \min \left[ \sum_{i=1}^D (C_{e,i} \cdot P_i + C_{R,i} \cdot R_i \cdot I \cdot (1-I) f) \right] \quad (5-1)$$

$$P_i + R_i \leq P_i^{\max} \quad (5-2)$$

$$P_i^{\min} \leq P_i \quad (5-3)$$

$$0 \leq R_i \leq R_i^{\max} \quad (5-4)$$

$$\sum P_i = P^{Load} \quad (5-5)$$

$$\sum R_i = R^{Load} \quad (5-6)$$

در (۵)،  $P_i^{\min}$  و  $P_i^{\max}$  به ترتیب حداکثر و حداقل توانایی تولید واحد  $i$ ام خواهد بود.  $R_i^{\max}$  نیز حداکثر توانایی تامین ظرفیت رزرو واحد  $i$ ام است. در این رابطه،  $P^{load}$  و  $R^{load}$  میزان بار مصرفی و ظرفیت رزرو مورد نیاز را مشخص می‌کند.

### ۳. الگوریتم یادگیری تقویتی

#### ۳.۱. الگوریتم یادگیری-Q

یادگیری تقویتی<sup>۸</sup> از یک نظریه روانشناختی حاصل از تجزیه و تحلیل رفتار حیوانات و سیستم‌های هوش مصنوعی الهام گرفته است. در یادگیری تقویتی عامل فرا می‌گیرد که در هر وضعیت برای حداکثر کردن سود خود در مدت طولانی چگونه عمل کند.

<sup>8</sup> Reinforcement Learning

۱. حداکثر کردن سود خود

۲. رضایت بخشی مورد نیاز از تولید روزانه

برای محاسبه استراتژی قیمت‌دهی در ابتدا نیاز به تعریف عناصر یادگیری تقویتی داریم [۹-۷]:

محیط: بازار در این استراتژی معرف محیط است که قیمت بسته شدن بازار را مشخص می‌کند.

عامل: هر کدام از تولیدکنندگان بیانگر عامل هستند که برای دستیابی به اهداف ذکر شده در بازار رقابت می‌کنند.

حالت: اطلاع هر یک از تولیدکنندگان در مورد قیمت بازار، حالت‌هایی از محیط را توصیف می‌کند. فضای حالت برای جلوگیری از نفرین ابعادی به تعداد محدودی حالت بین صفر و سقف قیمت بازار انرژی و رزرو تقسیم می‌شوند.

عمل: عملی که عامل انجام می‌دهد از قیمت پیشنهادی خود انتخاب می‌کند. از این رو فضای عمل قابل قبول به فواصل مساوی بین قیمت اولیه ارائه شده توسط تولیدکننده و سقف قیمت بازار تقسیم می‌شود. باید توجه داشت قیمت ارائه شده توسط تولیدکننده همواره از هزینه حاشیه‌ای تولید (هزینه تامین) بیشتر خواهد بود و در عرضه پله‌ای تولید برای هر پله، براساس قیمت پیشنهادی در آن پله، فضای عمل متفاوتی در نظر می‌گیریم.

پاداش: میزان سودی که عامل از شرکت در بازار در هر مرحله بدست می‌آورد، بیانگر پاداش است. برای سادگی آنالیز رفتار قیمت‌گذاری، تابع هزینه تولیدکننده را بصورت تابع خطی مانند (۱۰) در نظر می‌گیریم و تولیدکنندگان حداکثر ظرفیت خود را به بازار پیشنهاد می‌دهند.

$$Cost(E_i) = mcg \times P_i \quad (10)$$

$mc$  بیانگر قیمت حاشیه‌ای (هزینه تولید) هر تولیدکننده است.

تحت قانون قیمت‌گذاری  $PAB$  سود تولیدکننده از شرکت در بازار انرژی و رزرو به صورت (۱۱) محاسبه می‌شود [۶].

$$\pi_e = (C_{e,i} - mc_g).P_i + (C_{r,i} - mc_r).R_i . I(C_{r,i} - MCP_r) \quad (11)$$

که  $mc_r$  قیمت حاشیه‌ای ظرفیت رزرو است.

بعد از تعیین عناصر یادگیری تقویتی، مراحل قیمت دهی را می‌توان به شرح زیر انجام داد [۱۳]:

بهره برداری از پاداش های دریافتی را نشان می‌دهد.  $\gamma=0$  یعنی تنها پاداش آنی مرحله قبل را در نظر می‌گیرد و  $\gamma=1$  تمام پاداش ها را در طول زمان در نظر می‌گیرد [۱۳]. به بیان دیگر با تغییر مقدار  $\gamma$  می‌توان افق دید را تنظیم نمود.

$$a_t(s,a) = \frac{1}{\beta(s,a)} \quad (7)$$

### ۳.۲. SA-Q-Learning

الگوریتم  $SA-Q-Learning$  یک شبیه سازی ابتکاری ساده و اثربخش در حل مسائل بهینه‌سازی است. هدف این الگوریتم تحقیق در مورد کنترل تعادل میان کاوش و بهره‌برداری در طول تکامل الگوریتم یادگیری-Q است. این الگوریتم معیاری است که عامل برای انتخاب عمل بعدی خود استفاده می‌کند.

مراحل انتخاب عمل در  $SA-Q-Learning$  [۱۰]:

۱. عمل  $a_r$  را بطور تصادفی از تابع عمل هایی که می‌توانیم انجام دهیم، انتخاب می‌کنیم.

۲. عمل  $a_p$  را با توجه به خط مشی و سیاستی که از یادگیری قبل بدست آوردیم، محاسبه می‌کنیم.

$$a_p = \arg \max_a Q(s_{t+1}, a) \quad (8)$$

۳. تولید عدد تصادفی  $\xi$  در بازه  $(0,1)$

۴. اجرا  $a_n \in A$ :

$$a_n = \begin{cases} a_n & \text{if } \xi \geq \exp\left[\frac{Q(s_t, a_r) - Q(s_t, a_p)}{\text{Temperature}}\right] \\ a_r & \text{Otherwise} \end{cases} \quad (9)$$

$T$  دمای سیستم است که درجه تصادفی بودن حرکت به سوی جواب را تعیین می‌کند. اگر  $T_t$  دما در مرحله  $t$ ام تکرار باشد، بنابراین  $T_t = \beta T_{t-1}$  است که  $t$  یک عدد طبیعی و  $\beta$  یک مقدار ثابت نزدیک به ۱ برای اطمینان در سرد شدن (کاهش) تابع در نظر می‌گیریم [۱۳].

### ۴. استراتژی پیشنهادی قیمت‌دهی:

در این مقاله استراتژی قیمت‌دهی تولیدکنندگان با تاثیر عرضه پله‌ای تولید بر اساس الگوریتم  $SA-Q-Learning$  پیشنهاد شده‌است. در طول فرآیند قیمت‌دهی، تولیدکننده برای رسیدن به اهداف زیر تلاش می‌کند [۱۳]:

<sup>9</sup> Marginal cost

۱. یادگیری اولیه: مقدار دهی اولیه  $Q$  در جدول حالت-عمل

۲. قیمت دهی اصلی:

سقف قیمت در بازار انرژی و رزرو به ترتیب  $70\$/MWh$  و  $20\$/MWh$  در نظر گرفته شده است. همانطور که بیان شد فضای عمل قابل قبول به فواصل مساوی بین قیمت اولیه ارائه شده توسط تولیدکننده و سقف قیمت بازار انرژی و رزرو تقسیم می‌شود. فضای حالت نیز بین صفر و سقف قیمت بازار انرژی و رزرو تقسیم می‌شود. مقدار نرخ بهره‌برداری از پاداش‌های دریافتی برای همه واحدها برابر  $0.1$  و مقادیر  $t_0$  و  $\beta$  مورد استفاده در الگوریتم  $SA-Q-Learning$  را برابر  $100000$  و  $0.9$  در نظر می‌گیریم. از طرفی مقدار نرخ یادگیری را در مرحله یادگیری از رابطه (۷) و در مرحله قیمت‌دهی اصلی برابر  $0.9$  قرار می‌دهیم.

در ابتدا فرآیند قیمت‌دهی، یادگیری اولیه را برای بدست آوردن مقادیر اولیه مناسب برای جدول حالت-عمل برای  $100000$  ساعت مبادلاتی تحت هر دو سناریو به طور جداگانه اجرا می‌نماییم. بعد از تکمیل مرحله یادگیری، مقادیر اولیه جدول حالت-عمل برای مرحله قیمت دهی اصلی مورد استفاده قرار می‌گیرد. در نهایت الگوریتم قیمت دهی عامل‌ها در محیط بازار تحت دو سناریو مورد نظر اجرا می‌شود.

جدول (۱): مشخصات واحدهای تولیدی در سناریو اول

unit	$P_i^{max}$ (MW)	$P_i$ (MW)	$R_i$ (MW)	$MC_g$ (\$/MWh)	$MC_r$ (\$/MWh)
1	300	255	45	20	5
2	300	255	45	25	5
3	200	170	30	35	2.5

جدول (۲): مشخصات واحدهای تولیدی در سناریو دوم

unit	$P_i$ (MW)	$P_{i-L}$ (MW)	$MC_g$ (\$/MWh)	$R_{i-L}$ (MW)	$MC_r$ (\$/MWh)
1	255	155	20	45	5
		70	22.5		
		30	25		
2	255	190	25	30	7.5
		65	27.5	15	10
3	170	30	35	20	2.5
				12	5
				3	7.5

- ارسال پیشنهادات اولیه و تعیین حالت اولیه
- هر عامل براساس سیاست خود (رابطه (۸)) بهترین عمل را انتخاب می‌کند.
- ISO تمامی پیشنهادات را جمع آوری کرده و پخش بار را برنامه ریزی می‌کند.
- سود حاصل از مبادلات ساعتی بازار را دریافت می‌کند.
- در نهایت مقدار  $Q$  برای هر عامل با توجه به معادله (۶) بروزرسانی می‌شود.

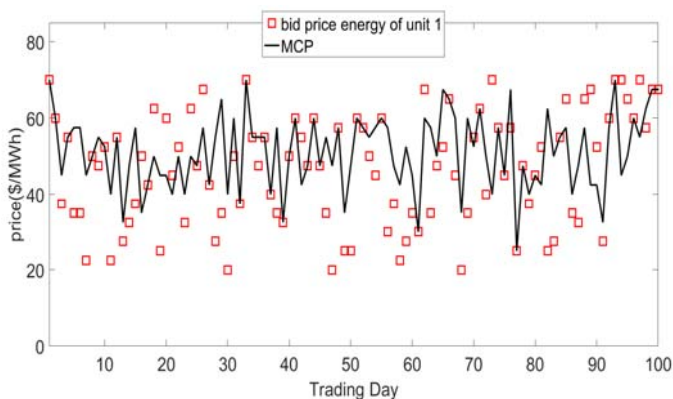
در روش پیشنهادی، تولیدکنندگان برای کسب سود بیشتر و هوشمندانه‌تر از روش پله‌ای برای عرضه تولید خود در بازار استفاده می‌کنند. در قیمت‌دهی پله‌ای تولیدکننده ظرفیت تولید خود را در مقادیر و قیمت‌های مختلف پیشنهاد می‌دهد. در این حالت، پله‌های انتهایی ظرفیت خود را نسبتاً کوچک در نظر می‌گیرند، به نحوی که برنده شدن یا نشدن این مقدار ظرفیت در بازار، در درآمدش تاثیر چندانی نداشته باشد و فقط برای شناسایی وضعیت بازار و پیش بینی نسبی قیمت رقبا از این پله استفاده می‌کند.

## ۵. مطالعات عددی:

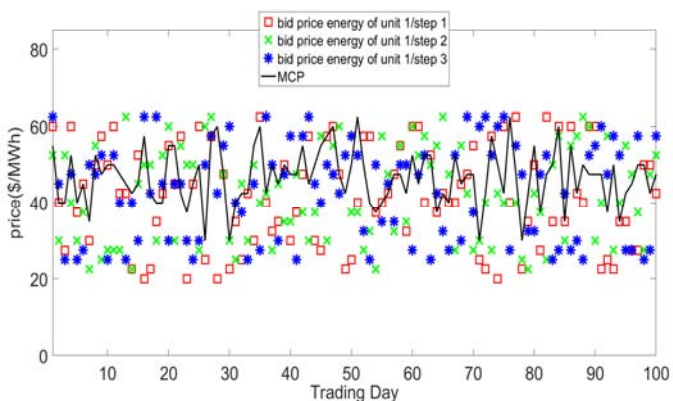
قصد داریم در این بخش سیستم تستی را تحت دو سناریو متفاوت مورد بررسی قرار دهیم. در سناریو اول، قیمت‌دهی و سود واحدهای تولیدی در بازار انرژی و رزرو در حالتی که تولید کننده تنها از یک پله برای پیشنهاد قیمت خود استفاده می‌کند و در سناریو دوم تولیدکنندگان ظرفیت تولیدی خود را به چند پله تقسیم کرده، سپس برای هر پله بصورت صعودی اقدام به قیمت‌دهی می‌کند.

مشخصات سیستم تست مورد مطالعه در دو سناریو، شامل ۳ تولیدکننده در جدول (۱) و (۲) ارائه شده است. شرکت کنندگان بازار حداکثر ظرفیت خود را به بازار پیشنهاد می‌دهند. هر تولیدکننده ۱۵٪ حداکثر ظرفیت تولیدی خود را برای بازار رزرو و مابقی را برای بازار انرژی در نظر می‌گیرد. بار پیش بینی شده توسط ISO برای ۲۴ ساعت روزانه در شکل (۲) نمایش داده شده است. همچنین میزان ظرفیت رزرو مورد نیاز برای هر ساعت ۱۵٪ بار پیش‌بینی شده ISO است.

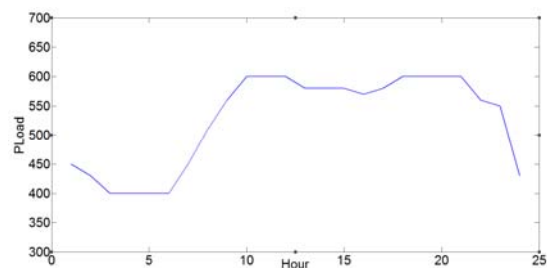
قیمت بسته شدن بازار رزرو و همچنین پیشنهاد قیمت تولیدکننده ۳ در بازار رزرو، در کم باری (ساعت ۳) تحت دو سناریو مورد نظر نمایش می‌دهند. در شکل (۵)، تولیدکننده ۳ که نسبت به رقبا دارای هزینه تامین رزرو کمتری است سعی خود را در برنده شدن در بازار رزرو معطوف کرده و اکثر مواقع قیمتی کمتر از قیمت تسویه بازار پیشنهاد داده و برنده بازار بوده است. در سناریو دوم ظرفیت رزرو خود را در سه پله با قیمت بیشتر ارائه می‌دهد تا شانس خود را برای کسب سود بیشتر در این بازار امتحان کند. در شکل (۶) مشاهده می‌شود، که پله‌های اول و دوم که سهم بیشتری از ظرفیت رزرو را دارند، برنده بازار رزرو خواهند بود. همچنین پیشنهاد قیمت پله‌های سوم اغلب بیشتر از قیمت تسویه بازار است و شانس کمی برای برنده شدن در بازار دارد، در شکل (۶) بیانگر این مفهوم است.



شکل (۳): قیمت بسته شدن بازار انرژی و همچنین پیشنهاد قیمت تولیدکننده ۱ در ساعت ۳ تحت سناریو اول به مدت ۱۰۰ روز



شکل (۴): قیمت بسته شدن بازار انرژی و همچنین پیشنهاد قیمت تولیدکننده ۱ در ساعت ۳ تحت سناریو دوم به مدت ۱۰۰ روز



شکل (۵): بار پیش بینی شده توسط ISO برای ۲۴ ساعت

جهت بررسی تاثیر عرضه پله‌ای تولید در استراتژی قیمت‌دهی، تولیدکننده‌های ۱ و ۳ را به عنوان دو تولیدکننده که هزینه تولید انرژی و هزینه رزرو کمتری نسبت به سایر رقبا دارند، در نظر می‌گیریم. تحت دو سناریو بیان شده، قیمت‌دهی و سود تولیدکننده ۱ را در بازار انرژی و قیمت‌دهی و سود تولیدکننده ۳ را در بازار رزرو مورد بررسی قرار خواهیم داد.

تولیدکننده ۱ در سناریو اول تمام ظرفیت تولید انرژی خود را در یک پله در بازار انرژی پیشنهاد می‌دهد و در سناریو دوم این مقدار تولید را در سه پله با قیمت‌های مختلف بصورت صعودی پیشنهاد می‌دهد. شکل‌های (۳) و (۴) قیمت بسته شدن بازار انرژی و همچنین پیشنهاد قیمت تولیدکننده ۱ در بازار انرژی، در کم باری (ساعت ۳) تحت دو سناریو مورد نظر نمایش می‌دهند. همانطور که در شکل (۳) مشاهده می‌شود، تولیدکننده ۱ که نسبت به بقیه رقبا دارای هزینه تولید انرژی پایین‌تری است، با ارائه پیشنهاد پایین‌تر از قیمت تسویه بازار اغلب برنده بازار خواهد بود. این تولیدکننده که به برنده شدن خود در بازار انرژی اطمینان دارد، برای کسب سود بیشتر، تحت سناریو دوم، تولید خود را در سه پله با قیمت‌های مختلف ارائه می‌دهد. با توجه به شکل (۴) پله‌های اول و دوم به دلیل ارائه قیمت کمتر نسبت به قیمت پیشنهادی دو تولیدکننده دیگر، برنده بازار خواهند بود و اغلب پیشنهادی کمتر از قیمت تسویه بازار می‌دهد. از طرفی نقاط پله سوم (step 3 در شکل (۴))، که با قیمت بیشتری ارائه شده، شانس کمتری برای برنده شدن در بازار دارد و قیمت‌های پیشنهادی آن اغلب برابر یا بالاتر از قیمت تسویه بازار است. اما از آنجا که مقدار ظرفیت این پله کوچک اتخاذ می‌شود، برنده شدن یا نشدن این پله در بازار انرژی، در درآمد واحد تولیدی تاثیر چندانی ندارد و فقط جهت بالا بردن قیمت پیشنهادی سایر رقبا در راستا برنده شدن خود در بازار است.

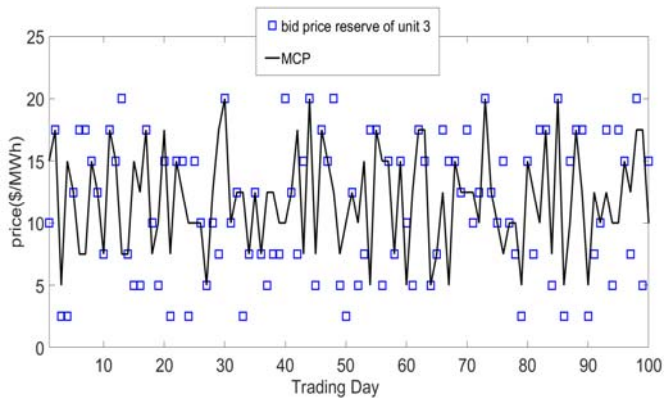
تولیدکننده ۳ نیز در سناریو اول تمام ظرفیت رزرو خود را در یک پله در بازار رزرو پیشنهاد می‌دهد و در سناریو دوم این ظرفیت را در سه پله با قیمت‌های مختلف بصورت صعودی پیشنهاد می‌دهد. شکل‌های (۵) و (۶)

## ۶. نتیجه گیری

در این مقاله، الگوریتم یادگیری-Q جهت شبیه‌سازی همزمان بازار انرژی و رزرو چرخشی پیشنهاد شد که تاثیر عرضه پله‌ای تولید بر رفتار شرکت‌کنندگان در مناقصات بازار مورد تجزیه و تحلیل قرار گرفته است. اجرای مناقصات بازارهای انرژی و رزرو در سیستم مورد مطالعه بصورت بازار روز پیش‌رو و بر اساس قانون تسویه پرداخت براساس پیشنهاد در نظر گرفته شده است. نتایج شبیه سازی بر روی سیستم نمونه تحت دو سناریو مختلف نشان می‌دهند که با در نظر گرفتن پله‌ای تولید، تولید کننده توانایی کسب سود بیشتر در مناقصات بازار را دارد. علاوه بر موارد فوق در الگوریتم یادگیری تقویتی ترکیبی از دو روش یادگیری-Q و روش سرد شدن تدریجی برای این مساله پیشنهاد گردید که این نوع یادگیری نیز در ماکزیمم کردن سود حاصل هم در تاثیر داشت.

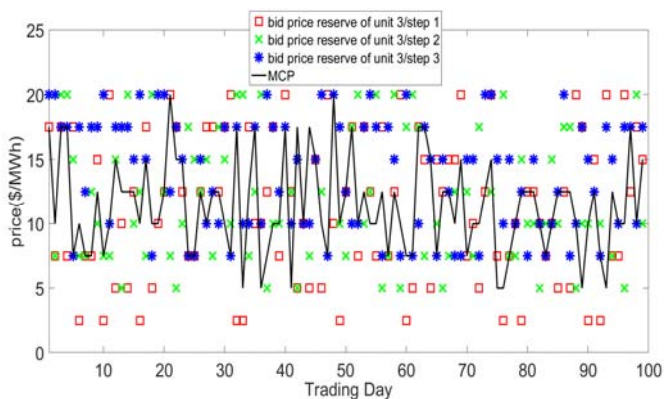
## منابع

- [1] S. S. Oren, "Design of ancillary service markets," Proceedings of the 34th Annual Hawaii International Conference on System Sciences, 2001.
- [2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
- [3] Jianxue Wang; Xifan Wang; Yang Wu, "Operating reserve model in the power market," IEEE Transactions on Power Systems, vol. 20, pp. 223-229, 2005.
- [4] A. Ahmadi-Khatir; R. Cherkaoui, "A probabilistic joint energy and spinning reserve market model," IEEE PES General Meeting, 2010.
- [5] Inga Umbrasko; Renata Varfolomejeva; Anatolijs Mahnitko, "Modeling of the generating company behavior in energy and reserve market," 11th International Conference on Environment and Electrical Engineering (EEEIC), pp. 1070-1074, 2012.
- [6] J. Khorasani; H. Rajabi Mashhadi, "Bidding analysis in joint energy and spinning reserve markets based on pay-as-bid pricing," Transmission & Distribution IET Generation, vol. 6, pp. 79-87, 2012.
- [7] Gaofeng Xion, T. Hashiyama and S. Okuma, "An electricity supplier bidding strategy through Q-Learning", IEEE Power Engineering Society Summer Meeting, Vol. 3, pp. 1516-1521, 2002.
- [8] Anastasios G. Bakirtzis, Athina C. Tellidou, "Agent-Based Simulation of Power Markets under Uniform and Pay-as-Bid Pricing Rules using Reinforcement Learning", IEEE PES Power Systems Conference and Exposition, Vol. 3, pp. 1168-1173, 2006.
- [9] G. Xiong, S. Okuma and H. Fujita, "Multi-agent Based Experiments on Uniform Price and Pay-as-Bid Electricity Auction Markets", IEEE International Conference on Electric Utility Deregulation, Restructuring and Power Technologies, Vol. 1, pp. 72-76, 2004.
- [10] Thanhquy Bach; Jiangang Yao; Jia Wang; Yang Shengjie, "Research and Application of the Q - Learning for Wholesale Power Markets", International Conference on Consumer Electronics, Communications and Networks (CECNet), pp. 1192-1197, 2012.
- [11] Gan, D. and Litvinov, E., "Energy and Reserve Market Designs With Explicit Consideration to Lost Opportunity Costs", IEEE Transactions on Power Systems, Vol. 18, No. 1, pp. 53-59, February 2003.
- [12] Chen, J., et al, "Coordinated Interchange Scheduling and Opportunity Cost Payment: A Market Proposal to Seams Issues", Proceedings of the 37th Hawaii International Conference on System Science, 2004.
- [13] Zhigang Liao, Ly-Fie Sugianto, "Using Q-learning to model bidding behaviour in electricity market simulation", IEEE Symposium on Computational Intelligence in Multicriteria Decision-Making, 2011.



شکل شماره (۵): قیمت بسته شدن بازار رزرو و همچنین پیشنهاد قیمت

تولیدکننده ۳ در ساعت ۳ تحت سناریو اول به مدت ۱۰۰ روز



شکل شماره (۶): قیمت بسته شدن بازار رزرو و همچنین پیشنهاد قیمت

تولیدکننده ۳ در ساعت ۳ تحت سناریو دوم به مدت ۱۰۰ روز

جدول (۳): سود پیش‌بینی شده در بازار روز پیش‌رو

Unit	سناریو اول (\$)	سناریو دوم (\$)
سود تولیدکننده ۱ در بازار انرژی روز پیش‌رو	221790	240825
سود تولیدکننده ۳ در بازار رزرو روز پیش‌رو	46788	50550

سود به‌دست آمده در دوره ۲۴ ساعت بازار روز پیش‌رو در جدول (۳)

تحت هر دو سناریو و در بازارهای مربوطه ارائه شده است. همانطور که مشاهده می‌شود تولیدکننده ۱ با عرضه پله‌ای تولید خود در بازار انرژی و تنها با فروش بخشی از ظرفیت خود، سود بیشتری نسب به حالتی که تمام ظرفیت خود را در یک پله ارائه می‌دهد کسب می‌کند. تولیدکننده ۳ نیز با ارائه پله ای ظرفیت رزرو خود در بازار رزرو پیش‌رو سود بیشتری کسب خواهد کرد.