بسم الله الرحمن الرحیم

تقدیم به

پدر و مادرم

و

دو خواهرم

که همیشه بهترین اند...

سپاس‌گزاری

حمد و سپاس بی‌نیاز بنده‌نواز را که یاریگر هر لحظه بود و به ما نعمت بخشید. ستایش از آن اوست و هر آنکس که در رسالت آموزگاری آموخت که آموختن برترین اندیشه است.

حال که به یاری پروردگار متعال این تحقیق به پایان رسیده، شایسته است از همه‌ي عزیزانی که اینجانب را در این راه مورد عنایت خود قرار داده‌اند تشکر و قدردانی نمایم.

بدین‌وسیله از استاد محترم راهنما جناب آقاي دکتر بادامچی‌زاده که در طول تدوین و ارزیابی پایان‌نامه با بصیرت و سعه‌ي‌صدر خود اینجانب را راهنمایی نمودند تشکر و قدردانی می‌نمایم. همچنین از زحمات فراوانی که استاد محترم مشاور جناب آقای دکتر یاری در طی این پایان‌نامه متقبل شدند تشکر و از جدیت ایشان در بررسی و ارزیابی مراحل مختلف این تحقیق سپاسگذاری می‌نمایم. بدون شک نمی‌توان راهنمایی‌هاي ارزنده‌ي استاد محترم داور جناب آقای دکتر پاشا و نماینده‌ي تحصیلات تکمیلی جناب آقای دکتر صالحی‌راد را نادیده گرفت و از زحمات ایشان نیز قدردانی می‌شود.

فهرست مطالب

فهرست شکل‌ها

چکیده

در این پایان‌نامه سرویس‌دهنده‌ی ناپایا و گذرا در صف‌هاي $M/G/1$ و $M^{[X]}/G/1$ با شیوه‌ي (p,N) - کنترل، بررسی می‌شود. سرویس‌دهنده براساس فرایند پواسون خراب شده و زمان‌هاي تعمیر دارای توزیع دلخواه هستند. از روش ماکزیمم‌آنتروپی براي به‌دست آوردن توزیع‌هاي احتمال تعداد متقاضیان و میانگین زمان انتظار در سامانه استفاده می‌کنیم. تحلیل نسبی میان نتایج تقریبی به‌دست آمده از روش ماکزیمم‌آنتروپی و نتایج دقیق برای سه توزیع مختلف زمان‌هاي سرویس و تعمیر گرفته است. نتایج عددی نشان می‌دهند که روش ماکزیمم‌آنتروپی به اندازه‌ي کافی برای کارهای عملی دقیق هستند.

کلید واژه: روش لاگرانژ، شیوه‌ي (p,N)– کنترل، ماکزیمم‌آنتروپی، سرویس‌دهنده‌ی ناپایا، صف $M/G/1$، صف $M^{[X]}/G/1$.

فصل 1

مقدمه‌ای بر نظریه‌ي صف‌بندي

1.1 مقدمه

همه‌ی ما ناراحتی انتظارکشیدن در صف را تجربه کرده‌ایم. متأسفانه این پدیده با افزایش جمعیت و شهری‌شدن روزافزون جامعه بیش‌ازپیش گسترش می‌یابد. یک متقاضی اینگونه انتظارکشیدن را دوست ندارد و مدیران مؤسسات نیز انتظارکشیدن متقاضیان را دوست ندارند زیرا ممکن است این صف‌ها برای آنان هزینه‌هایی داشته باشد. براي دانستن زمان سرویس مناسب نیاز به دانستن پاسخ برخی سوال‌ها است: «چه مدت متقاضی باید انتظار بکشد؟» و «چند متقاضی در صف خواهند بود؟» نظریه‌ي صف‌بندی سعی دارد به این قبیل سوال‌ها براساس تحلیل توزیع‌هاي احتمال پاسخ دهد. مدل‌هاي صف‌بندی یکی از شاخه‌های فرایندهای تصادفی است که هم از لحاظ کاربردی و هم از لحاظ نظري مورد علاقه‌ی بسیاری از محققین آمار می‌باشد. این مدل‌ها به بررسی سامانه‌هایی می‌پردازد که در آن‌ها خدماتی را به متقاضیان ارائه می‌دهند. یکی از هدف‌هاي مهم این مدل‌ها، چگونگی سرویس‌دهی به متقاضیان است به طوری‌که این سرویس به شکل بهینه ارائه گردد.

یک سامانه‌ي صف‌بندی را می‌توان چنین توصیف کرد که متقاضیان برای اخذ سرویس مراجعه می‌کنند اگر ارائه‌ی سرویس بلافاصله مقدور نباشد منتظر می‌مانند و بعد از اخذ سرویس سامانه را ترک می‌کنند. در اینجا باید توجه داشت که اصطلاح متقاضی به معنای کلی آن به‌کار رفته است و لزوما به معنای یک انسان نیست. مثلا یک متقاضی می‌تواند بلبرینگی باشد که در خط تولید برای صیقل‌خوردن قرار گرفته است، هواپیمایی باشد که در خط پرواز برای بلندشدن از زمین انتظار می‌کشد و یا برنامه‌اي کامپیوتری باشد که برای اجرا در انتظار است.

هر سامانه‌ي صف‌بندي داراي شش مشخصه‌ی پایه‌اي است که عبارتند از: الگوي ورود متقاضیان، الگوی ارائه‌ی سرویس توسط سرویس‌دهنده‌ها، نظم صف، ظرفیت سامانه، تعداد باجه‌هاي سرویس و تعداد مراحل سرویس [1].

2.1 تعاریف و مفاهیم

در این پایان‌نامه مقدمات نظریه‌ی صف دانسته فرض شده و فقط به معرفی برخی از مفاهیم پایه‌اي بسنده می‌شود. برای مطالب بیشتر به [1] مراجعه شود.

1.2.1 اندازه‌هاي مؤثربودن

هر مدل صف‌بندي داراي مشخصه‌هایی است که به کمک آن‌ها می‌توان مدل صف‌بندي را تحلیل و بهینه کرد. این مشخصه‌ها را اندازه‌های مؤثربودن می‌نامیم. بهینه‌سازي در سامانه‌هاي صف‌بندي چه از نظر سودآوري و چه از نظر کاهش زمان انتظار متقاضیان، با توجه به توابع هزینه‌ي مفروض و براساس اندازه‌هاي مؤثربودن مانند میانگین تعداد در سامانه، میانگین زمان انتظار متقاضیان در سامانه، میانگین دوره‌ي اشتغال (و یا دوره‌ی بیکاری) سرویس‌دهنده‌ها و ... انجام می‌گیرد. براي به‌دست آوردن این اندازه‌ها لازم است متغیرهای موجود در مدل شناسایی شده و توزیع‌هاي آن‌ها به‌دست آورده شوند. اما این توزیع‌ها دارای پارامترهایی هستند که معمولا نامعلوم می‌باشند. بنابراین برای بهینه‌سازي براساس اندازه‌هاي مؤثربودن نیازمند براورد این پارامترها هستیم.

برای براورد این پارمترها از روش‌های کلاسیک و یا دیگر روش‌ها استفاده می‌شود. در دیگر روش‌ها مانند روش بیز، پارامترهاي مدل خود به‌عنوان متغیرهای تصادفی هستند و آماردان با توجه به تجربیات و نظر شخصی خود باید یک تابع چگالی پیشین برای پارامترهای مدل انتخاب کند و براساس این تابع چگالی و همچنین نمونه‌ای که از جامعه به‌دست آورده است تابع چگالی پسین را محاسبه نماید.

معمولا فرایندهای ورود یا سرویس تصادفی‌اند، لذا اندازه‌های مؤثربودن برای بررسی سامانه، یعنی مقادیری مانند زمان انتظار، تعداد متقاضیان در سامانه، طول دوره‌ی اشتغال یا تعطیلی سرویس‌دهنده‌ها نیز متغیرهای تصادفی خواهند بود و تعیین توزیع احتمال این متغیرهای تصادفی یا حداقل مقادیر امیدریاضی آن‌ها مدنظر است.

کار تحلیل‌گر صف‌بندی یکی از دو مورد زیر است:

- تعیین اندازه‌هاي مؤثربودن فرایند داده‌شده
- طرح یک سامانه‌ي بهینه

براي انجام مورد دوم، باید زمان‌هاي انتظار، طول صف و نظیر این‌ها را با خواص مفروض جریان ورودی و شیوه‌های سرویس مرتبط کند. از طرف دیگر، برای طرح یک سامانه، تحلیل‌گر احتمالا خواستار تعدیل زمان انتظار متقاضی در برابر زمان فراغت سرویس‌دهندگان برحسب ساختار ذاتی هزینه‌ها است. اگر هزینه‌ي انتظار و فراغت سرویس‌دهنده را بتوان مستقیما به‌دست آورد، از آن‌ها می‌توان برای تعیین تعداد بهینه‌ي باجه‌های فعال و تعیین نرخ راه‌اندازي این باجه‌ها استفاده کرد. همچنین، برای طرح مکان انتظار، لازم است اطلاعاتی درباره‌ي حجم ممکن صف براي تدارک مکان انتظار داشته باشیم. ممکن است هزینه‌ي مکان نیز مطرح باشد که باید این هزینه همراه با هزینه‌هاي انتظار متقاضی و اوقات فراغت سرویس‌دهنده براي به‌دست آوردن طرح سامانه‌ي بهینه بررسی شوند. در هر حال، تحلیل‌گر کوشش می‌کند که به‌صورت تحلیلی مساله‌اش را حل کند. در صورت عدم توفیق، باید به شبیه‌سازي متوسل گردد.

2.2.1 شیوه‌ي $(0,N)$ - کنترل

هنگامی که سرویس‌دهنده پشت باجه‌ي سرویس است سرویس‌دهی به متقاضیان را تا زمانی که صف تهی شود، ادامه می‌دهد؛ پی از تهی شدن صف باجه‌ي سرویس را می‌بندد و فقط زمانی برمی‌گردد که یا صفی به اندازه‌ي N ایجاد شده باشد و یا این‌که مجموع زمان انتظار متقاضیان در صف از سطح Q بیشتر گردد که Q یک مقدار مثبت ثابت و N یک عدد صحیح مثبت است. به‌عبارت دیگر سرویس فعال می‌شود هنگامی که اندازه‌ي صف N گردد و تعطیل است هنگامی که اندازه‌ي صف صفر است. این قاعده‌ی سرویس‌دهی با نماد $(0,N,Q)$ نشان داده می‌شود.

3.2.1 قانون لیتل

قانون لیتل رابطه‌اي بین میانگین تعداد متقاضیان در سامانه، میانگین نرخ ورود و میانگین زمان پاسخ به متقاضی در حالت پایا برقرار می‌کند.

فرض کنید $N(t)$ نشان‌دهنده‌ي تعداد متقاضیان در سامانه در زمان t، $A(t)$ معرف تعداد متقاضیان وارد شده به سامانه در فاصله‌ي زمانی $[0,t]$، $D(t)$ بیان‌گر تعداد متقاضیانی که در فاصله‌ي زمانی $[0,t]$ سامانه را ترک می‌کنند و T_i زمان پاسخ i- امین متقاضی باشد. واضح است که $N(t)=A(t)-D(t)$ (با این فرض که سامانه در زمان $t=0$ خالی است). ارتباط میان $A(t)$ و $D(t)$ در قسمت بالاي شکل 1.1 نشان داده شده است.

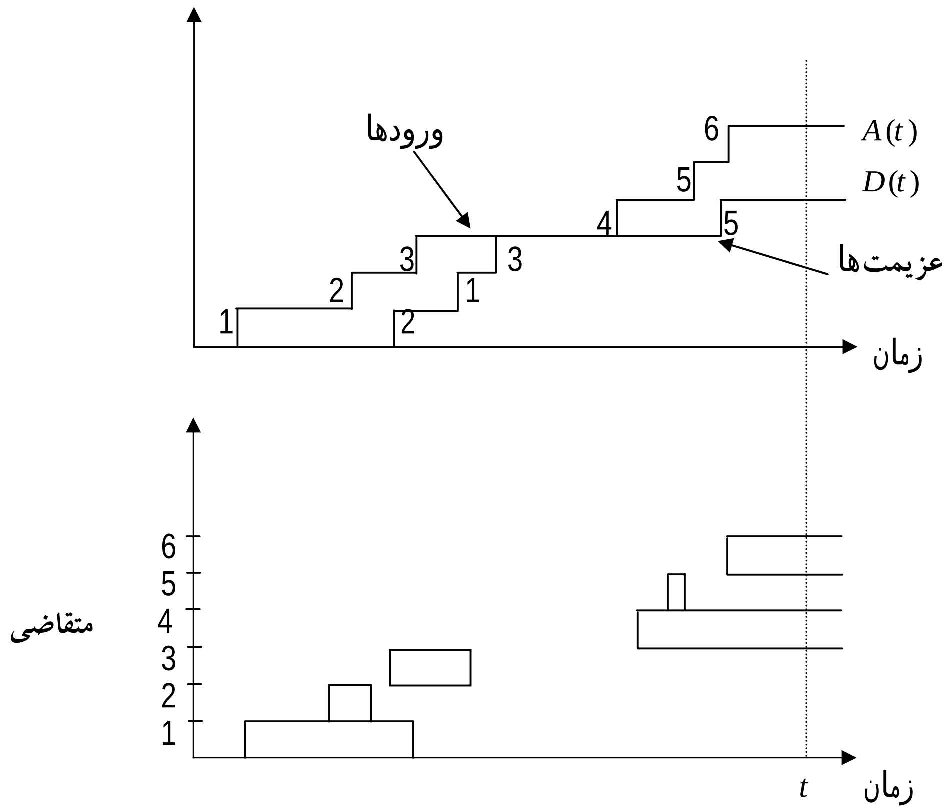
لازم به یاداوری است که ترتیب ورود و خروج متقاضیان الزاما یکی نیست. میانگین تعداد ورودي‌ها در فاصله‌ي زمانی $[0,t]$ برابر است با:

$$\overline{A}(t)=\frac{A(t)}{t} \tag{1.2.1}$$

و فرض می‌کنیم $\lambda=\lim_{t\to\infty}\overline{A}(t)$ موجود بوده و کراندار است. مقدار λ در دراز مدت به عنوان نرخ ورود در نظر گرفته می‌شود.

میانگین زمانی تعداد متقاضیان در سامانه برابر است با:

$$\overline{N}(t) = \frac{N(t)}{t} \tag{2.2.1}$$



شکل 1.1: قانون لیتل

در این‌جا نیز فرض می‌کنیم $\overline{N} = \lim_{t\to\infty}\overline{N}(t)$ وجود داشته و کراندار است. به‌طور مشابه میانگین زمان پاسخ به متقاضی را می‌توان به‌صورت زیر نوشت:

$$\overline{T}(t) = \frac{1}{A(t)}\sum_{i=1}^{A(t)} T_i \tag{3.2.1}$$

حال قسمت بالاي شکل 1.1 را در نظر بگیرید. ملاحظه می‌شود که همیشه $A(t) \geq D(t)$ و بنابراین $N(t) \geq 0$. مساحت بین دو منحنی برابر است با:

$$F(t) = \int_0^t \big(A(u) - D(u)\big)du = \int_0^t N(u)du \tag{4.2.1}$$

می‌توان $F(t)$ را به‌این صورت بیان کرد که نشان‌دهنده‌ي مجموع زمان‌های پاسخ متقاضیانی است که تا زمان t در سامانه حضور دارند:

$$\sum_{i=1}^{A(t)} T_i$$

که شامل کل زمان پاسخ به متقاضیانی است که تا زمان t در سامانه هستند (در قسمت پایین شکل 1.1، هر خط معرف زمان پاسخ به متقاضی است). بنابراین $F(t)$ را می‌توان به این صورت نیز نوشت:

$$F(t) = \sum_{i=1}^{A(t)} T_i - E(t) \tag{5.2.1}$$

با این فرض که $E(t)$ نسبتا کوچک است.

با برابر هم قراردادن طرفین رابطه‌هاي (4.2.1) و (5.2.1) داریم:

$$\int_0^t N(u)du = \sum_{i=1}^{A(t)} T_i - E(t) \tag{6.2.1}$$

با ضرب طرفین رابطه‌ي (6.2.1) به $\frac{1}{t}$ نتیجه می‌شود:

$$\frac{1}{t}\int_0^t N(u)du = \frac{A(t)}{t}\frac{1}{A(t)}\sum_{i=1}^{A(t)} T_i - \frac{E(t)}{t} \tag{7.2.1}$$

حال با استفاده از تعریف‌های بالا و گرفتن حد و این نکته که $\lim_{t\to\infty}\frac{E(t)}{t} = 0$، قانون لیتل را به‌صورت زیر داریم:

$$\overline{N} = \lambda \overline{T} \tag{8.2.1}$$

یا به‌طور معادل:

$$E(N) = \lambda E(T) \tag{9.2.1}$$

که $\overline{N}$ نشان‌دهنده‌ي میانگین تعداد متقاضیان در سامانه می‌باشد. صورت مشابهی از قانون لیتل برای میانگین تعداد متقاضیان در صف برقرار است که به‌صورت $\overline{N}_q$ نشان داده می‌شود و در آن متغیر

تصادفی N_q به تعداد متقاضیان موجود در صف اشاره دارد. همچنین میانگین زمان انتظار در صف یعنی زمان بین ورود یک متقاضی و شروع سرویس‌دهی به او را با W_q^* نشان می‌دهیم. بنابراین قانون لیتل را در این حالت می‌توان به‌صورت زیر نوشت:

$$\overline{N}_q = \lambda W_q^* \tag{10.2.1}$$

یا به‌طور معادل:

$$E\left(N_q\right) = \lambda E\left(W_q\right) \tag{11.2.1}$$

که در آن W_q مدت زمان انتظار در صف است.

4.2.1 سرویس‌دهنده‌ي ناپایا

به معناي سرویس‌دهنده‌اي است که به‌صورت غیر قابل پیش‌بینی خراب می‌شود.

5.2.1 سرویس‌دهنده‌ي گذرا

در این نوع سامانه‌ها، روشن و خاموش‌شدن سرویس‌دهنده به تعداد متقاضیان در آن بستگی دارد.

6.2.1 نظم $FCFS$[۱]

به معناي ارائه‌ی سرویس به‌ترتیب ورود است؛ به‌عبارت دیگر هر متقاضی که زودتر وارد صف شده باشد، زودتر سرویس می‌گیرد.

7.2.1 زمان پاسخ به متقاضی

زمان بین ورود متقاضی به سامانه و ترك آن بعد از گرفتن سرویس.

8.2.1 تعطیلی

روند تعطیلی به این صورت است که به محض خالی‌شدن سامانه، سرویس‌دهنده سامانه را براي تعطیلی تصادفی به طول V ترک می‌کند. هنگامی که سرویس‌دهنده از تعطیلی برمی‌گردد، اگر اندازه‌ي

[1] First Come, First Served

سامانه بزرگتر یا مساوی عدد N باشد، سرویس‌دهی به متقاضیان منتظر را آغاز می‌کند، درغیر این‌صورت سرویس‌دهنده در سامانه منتظر می‌ماند تا این‌که اندازه‌ي سامانه به N برسد یا از آن بیشتر شود.

9.2.1 حالت پایا

یعنی وضعیت سامانه مستقل از وضعیت آغازی و زمان طی شده است.

10.2.1 زمان تکمیل

شامل زمان سرویس‌دهی به متقاضی و زمان تعمیر سرویس‌دهنده می‌باشد.

11.2.1 چرخه‌ي اشتغال

به‌صورت مجموع دوره‌ي اشتغال و دوره‌ی بیکاری میباشد.

3.1 الگوي صف

فرض کنید متقاضیان در گروه‌هایی وارد می‌شوند که توزیع احتمال این گروه‌ها براساس پواسون مرکب با نرخ λ است. همچنین فرض کنید X_k نشان‌دهنده‌ي تعداد متقاضیان متعلق به k - امین گروه وارد شده است، به‌طوري‌که به ازاي $k=1,2,...$ ، X_kها داراي تابع احتمال زیر هستند:

$$P(X_k = n) = X_n \quad , \quad n = 1, 2, ...$$

متقاضیان در قالب دسته‌هایی وارد سامانه شده و به‌ترتیب ورودشان سرویس‌دهی می‌شوند. سرویس‌دهنده در یک زمان فقط یک متقاضی را می‌تواند سرویس بدهد. زمان‌هاي سرویس متقاضیان، متغیرهاي تصادفی مستقل و هم‌توزیع S با تابع توزیع کلی $S(t)$ است. سرویس‌دهنده ممکن است در هر لحظه از زمان با نرخ پواسون α خراب شود. سرویس‌دهنده‌ي از کار افتاده فورا تعمیر می‌شود. زمان‌هاي تعمیر، متغیرهاي تصادفی مستقل و هم‌توزیع R با تابع توزیع کلی $R(t)$ است. اگرچه هیچ سرویسی در طول دوره‌ی تعمیر سرویس‌دهنده‌ي از کار افتاده انجام نمی‌گیرد، اما ورود متقاضیان

برطبق فرایند پواسون ادامه می‌یابد. در حالتی که سرویس‌دهنده در طول سرویس‌دهی به متقاضیان خراب گردد، به تعمیر فرستاده می‌شود و هر متقاضی که در حال سرویس‌دهی بود باید براي تکمیل‌شدن سرویس باقیمانده‌اش منتظر سرویس‌دهنده بماند. به محض این‌که سرویس‌دهنده تعمیر شد شروع به سرویس‌دهی به متقاضیان منتظر می‌کند تا زمانی که سامانه خالی گردد.

ساده‌ترین رده‌ي مسائل صف‌بندی از نظر مفهومی رده‌ای است که در آن توزیع‌هاي احتمال براي توصیف الگوهای مراجعه و سرویس ضروری نیستند. با این حال، واحدهاي ورودي در نقاط معینی از زمان اتفاق می‌افتند و زمان‌های سرویس مقادیر مشخص و ثابت هستند. هر مدل صف‌بندی که در این رده قرار گیرد قطعی نامیده می‌شود، زیرا در هر حال، هیچ توزیع احتمال همراه با مساله وجود ندارد.

حالت مقدماتی نرخ ثابت مراجعات به یک باجه را که دارای نرخ ثابت سرویس است در نظر بگیرید. این‌گونه ورودي‌ها که با فاصله‌ی زمانی منظم صورت می‌گیرند به ترتیب ورودشان ($FCFS$) سرویس‌دهی می‌شوند. همچنین فرض می‌کنیم در زمان $t = 0$، هیچ متقاضی در خط انتظار نبوده و باجه‌ی سرویس تهی است. اگر λ را تعداد مراجعات در واحد زمان تعریف کنیم، آنگاه $\frac{1}{\lambda}$ متوسط فاصله‌ی زمانی ثابت بین مراجعات متوالی است. همچنین وقتی که سرویس‌دهنده مشغول انجام وظیفه است، اگر μ نرخ سرویس برحسب میزان سرویس‌های تکمیل‌شده در واحد زمان باشد، آنگاه متوسط زمان ثابت سرویس $\frac{1}{\mu}$ است.

4.1 کاربرد

موارد متعددی برای کاربرد مدل‌هاي صف‌بندي وجود دارد. چند نمونه از این موارد را در اینجا می‌آوریم. اولین مورد که به آن اشاره می‌شود، نخستین زمینه‌ي کاربرد آن یعنی مکالمات تلفنی است. بعضی موارد برجسته‌ي دیگر عبارتند از: تعمیر ماشین، جایگاه‌های پلیس‌راه، ایستگاه‌های تاکسی، کنترل موجودی، بارگیری و تخلیه‌ی کشتی‌ها، برنامه‌ریزی خدمات درمانی بیماران در درمانگاه‌ها، جریان خط

تولید، فرودگاه‌ها و بنادر، برنامه‌نویسی، سامانه‌های تولید/فهرست کالا و اخیرا ارتباط دیجیتال و شبکه‌هاي رایانه‌اي است.

5.1 تاریخچه

شیوه‌ي N - کنترل اولین بار توسط یادین و نائور[1] (1963) معرفی شد. هِیمن (1968) اولین کسی بود که صف $M/G/1$ با شیوه‌ي N - کنترل را مطالعه کرد. یک تحقیق جامع راجع به تعطیلی مدل‌هاي صف‌بندي توسط دوشی (1986) انجام گرفته است. کلا (1989) بحث‌هایی جزیی درباره‌ي سامانه‌هاي صف‌بندي $M/G/1$ و $M^{[X]}/G/1$ ارائه کرد. لی و همکاران (1995) سامانه‌ي صف‌بندي $M/G/1$ با شیوه‌ي N - کنترل را به تفصیل بررسی کردند. فاینبرگ و کیم (1996) صف $M/G/1$ با شیوه‌ي (p,N) - کنترل و با سرویس‌دهنده‌ي ناپایا را مطالعه کردند. اخیرا نیز انواع سامانه‌هاي صف‌بندي با شیوه‌ي N - کنترل توسط ژانگ و همکاران (1997)، کریشنا و همکاران (1998)، که (2001) و بسیاری دیگر مطالعه شده است.

وانگ و که (2002) ویژگی‌های حالت پایا مانند میانگین تعداد متقاضیان، میانگین طول دوره‌هاي بیکاری، خرابی، تعمیر و تکمیل براي صف $M/G/1$ با شیوه‌ي N - کنترل و سرویس‌دهنده‌ي ناپایا را بررسی کردند. که (2003) روش بهینه‌ي قابل کنترل سامانه‌هاي صف‌بندي $M^{[X]}/G/1$ با خرابی‌هاي سرویس‌دهنده و زمان‌هاي تعطیلی را بررسی کرد. سه سال بعد لین و که (2006) شیوه‌اي متفاوت را براي سامانه‌ي صف‌بندي $M^{[X]}/G/1$ با سرویس‌دهنده‌ي ناپایا و تعطیلی‌های مکرر ارائه دادند.

مدل‌هاي صف‌بندی با خرابی‌هاي سرویس‌دهنده نشان‌دهنده‌ي حالت‌های واقعی هستند. سرویس‌دهنده می‌تواند در طول کار به‌صورت پیش‌بینی نشده خراب شود. گِی‌ور[2] (1962) سامانه‌ي صف‌بندي $M/G/1$ با خرابی سرور را مطرح کرد. سامانه‌ي گی‌ور توسط سِنگوپتا[3] (1990) به حالت $GI/G/1$ بسط داده شد. تانگ (1997) خرابی و تعمیر سرویس‌دهنده‌های ناپایا را بررسی کرد.

[1] YADIN and NAOR
[2] GAVER
[3] SENGUPTA