QUEUEING THEORY AND PROCESS FLOW PERFORMANCE

Chang-Sun Chin[1]

ABSTRACT

Queuing delay occurs when a number of entities arrive for services at a work station where a server(s) has limited capacity so that the entities must wait until the server becomes available. We see this phenomenon in the physical production environment as well as in the office environment (e.g., document processing). The obvious solution may be to increase the number of servers to increase capacity of the work station, but other options can attain the same level of performance improvement.

The study selects two different projects, investigates their submittal review/approval process and uses queuing theory to determine the major causes of long lead times. Queuing theory provides good categorical indices—variation factor, utilization factor and process time factor—for diagnosing the degree of performance degradation from queuing. By measuring the magnitude of these factors and adjusting their levels using various strategies, we can improve system performance. The study also explains what makes the submittal process of two projects perform differently and suggests options for improving performance in the context of queuing theory.

KEY WORDS

Process time, queueing theory, submittal, variation, utilization

INTRODUCTION

Part of becoming lean is eliminating all waste (or *muda* in Japanese). Waste is "any activity which consumes resources but creates no value" (Womack and Jones 2003). Waiting, one of seven wastes defined by Ohno of Toyota (Ohno 1988), can be seen from two different views: "work waiting" or "worker waiting." "Work waiting" occurs when servers (people or equipment) at work stations are not available when entities (jobs, materials, etc) arrive at the work stations, that is, when the servers are busy and entities wait in queue. "Worker waiting" occurs when servers at work stations are ready to serve, but entities are not available, that is, no jobs arrive at the work stations so servers are idle. However, it is clear that both cases consume resources without creating value; "work waiting" consumes space for entities to wait until the server is ready, and "worker waiting" consumes server's capability without actual production. This study explores the underlying causes of waiting in a process flow and finds improvement methods from the queuing perspective.

---

[1] Ph.D., Honorary Fellow, Construction Engineering and Management Program, Department of Civil and Environmental Engineering, University of Wisconsin, Madison, chin2@wisc.edu