

UC Irvine

ICS Technical Reports

Title

A Note on Response Time and Saturation

Permalink

<https://escholarship.org/uc/item/60f3n3hp>

Author

Tonge, Fred M.

Publication Date

1974

Peer reviewed

A NOTE ON RESPONSE TIME
AND SATURATION

Fred M. Tonge

TECHNICAL REPORT #40 - January 1974

A NOTE ON RESPONSE TIME AND SATURATION

A standard result (1,2,3) in finite population models of time-sharing systems, given appropriate assumptions as to the distribution of service requests and think time, is that average response time is given as:

$$T = \frac{M S}{1 - \Pi_0} - U$$

where M = number of consoles

S = average service request

U = average think time

Π_0 = probability that no consoles are receiving or
awaiting service

In this context, Scherr (1, pg.72) defines saturation as follows:

"Saturation occurs when the probability of zero users waiting for service is lower than some small number, ϵ ." Kleinrock (3, pg. D121) defines as the saturation point for an M -console system:

$$M^* = \frac{S + U}{S}$$

and states "This is similar to a definition given by Scherr". Scherr (1, pg. 108) also states that "...there is no relationship between the point at which the system saturates and whether or not the response times at that point are acceptable."

This note presents

1. a restatement of average response time as a multiplier of average service request;

2. a computational form of this restatement as a recursive relationship
3. the behavior of this multiplier at the saturation point.

I. Restatement of response time as a multiplier of average service request:

Let $R = S/U$,

$$\text{then } T = \frac{MS}{1-\Pi_0} - \frac{S}{R} = S \left[\frac{M}{1-\Pi_0} - \frac{1}{R} \right].$$

$$\text{And letting } Q(M) = \left[\frac{M}{1-\Pi_0} - \frac{1}{R} \right],$$

$$T = S \cdot Q(M).$$

Henceforth we are concerned with $Q(M)$.

$$\text{Note that } \Pi_0 \text{ is defined as } \left[\sum_{j=0}^M \frac{M!}{(M-j)!} R^j \right]^{-1}$$

Note also that $Q(M)$ is a function of R .

$$Q(M) = \frac{M}{1 - \frac{1}{R}} - \frac{1}{R} = \frac{M \sum_{j=0}^M \left[\frac{M!}{(M-j)!} R^j \right]}{M \sum_{j=0}^M \left[\frac{M!}{(M-j)!} R^j \right] - 1} - \frac{1}{R}$$

$$= \frac{M \sum_{j=0}^M \left[\frac{R^j}{(M-j)!} \right] - M \sum_{j=1}^M \left[\frac{R^{j-1}}{(M-j)!} \right]}{M \sum_{j=1}^M \left[\frac{R^j}{(M-j)!} \right]} = \frac{M \sum_{j=1}^M \left[\frac{R^j}{(M-j)!} \right] - M \sum_{j=2}^M \left[\frac{R^{j-1}}{(M-j)!} \right]}{M \sum_{j=1}^M \left[\frac{R^j}{(M-j)!} \right]}$$

$$= \frac{M \sum_{j=1}^{M-1} \left[\frac{R^j}{(M-j)!} \right] + MR^M \sum_{j=1}^{M-1} \left[\frac{R^j}{(M-1-j)!} \right]}{M \sum_{j=1}^M \left[\frac{R^j}{(M-j)!} \right]} = \frac{M \sum_{j=1}^M \left[\frac{j R^j}{(M-j)!} \right]}{M \sum_{j=1}^M \left[\frac{R^j}{(M-j)!} \right]}$$

$$= \frac{\sum_{j=1}^M j L(M,j)}{\sum_{j=1}^M L(M,j)}, \text{ where } L(M,j) = \frac{R^j}{(M-j)!}$$

II. Computational Simplification of Q(M)

Note that: $L(M,j) = R \cdot L(M-1,j-1)$ for $M > 1$, $M \geq j > 1$, and

$$L(1,1) = R$$

$$\text{Let } Z(M) = \sum_{j=1}^M L(M,j)$$

$$\text{Also, } Z(M) = L(M,1) + \sum_{j=2}^M L(M,j) = \frac{R}{(M-1)!} + \sum_{j=2}^M R \cdot L(M-1,j-1)$$

$$= R \left[\frac{1}{(M-1)!} + \sum_{j=1}^{M-1} L(M-1,j) \right] = R \left[\frac{1}{(M-1)!} + Z(M-1) \right]$$

$$\text{Then } Q(M) = \frac{\sum_{j=1}^M jL(M,j)}{Z(M)}, \text{ and } \sum_{j=1}^M jL(M,j) = Q(M)Z(M)$$

$$Q(M) = \frac{L(M,1) + \sum_{j=2}^M jL(M,j)}{Z(M)} = \frac{L(M,1) + R \sum_{j=2}^M jL(M-1,j-1)}{Z(M)}$$

$$= \frac{L(M,1) + R \sum_{j=1}^{M-1} (1+j)L(M-1,j)}{Z(M)}$$

$$= \frac{\frac{R}{(M-1)!} + R \sum_{j=1}^{M-1} L(M-1,j) + R \sum_{j=1}^{M-1} jL(M-1,j)}{Z(M)}$$

$$= \frac{Z(M) + R \cdot Q(M-1)Z(M-1)}{Z(M)}$$

$$= 1 + R \frac{Z(M-1)}{Z(M)} Q(M-1)$$

III. Behavior of the average service time multiplier

[Q(M,R)] at saturation

The multiplier $Q(M)$ is graphed in Figure 1 as a function of M (number of consoles) for various values of R . As can be seen, the change in Q with a unit change in M approaches 1 beyond the saturation point.

Using Kleinrock's definition of saturation, the last console to be added before interference occurs on the average is given by

$$M^* = \frac{1}{R} + 1$$

The average service time multiplier $Q(M^*)$, is graphed as a function of M^* in Figure 2, as is $Q(M^*+1) - Q(M^*)$, the incremental increase in response time for the first additional user beyond saturation.

Note that, beyond approximately 15 terminals, this difference declines very slowly, and for most practical purposes can be considered a constant (~ 0.43).

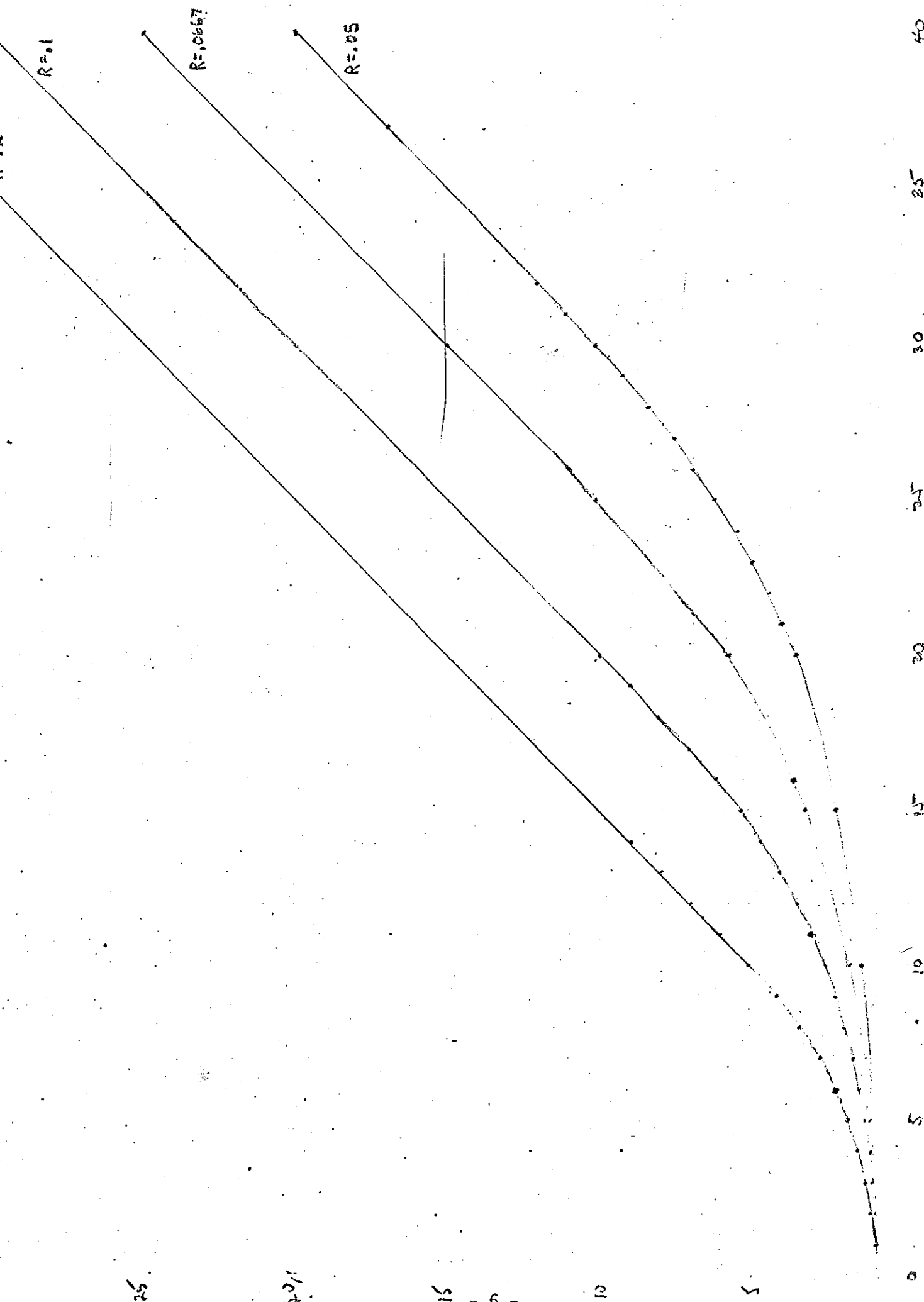


Figure 1

11/8

$$Q(M^*) \text{ for } M^* = \frac{1}{R} + 1$$

$$Q(M^*+1) - Q(M^*)$$

5.0

4.0

3.0

- 7 -

2.0

1.0

0

5

10

15

20

25

30

35

40

Figure 2

M*

References

1. Scherr, A.L., An Analysis of Time-Shared Computer Systems
M.I.T. Project MAC report MAC-TR-18 (Thesis) June 1965
2. Estrin, G. and Kleinrock, L., "Measures, Models, and Measurements
for Time-Shared Computer Utilities",
Proc. ACM 22nd National Meeting, August 1967
3. Kleinrock, L., "Certain Analytic Results for Time-Shared Processors",
Proc. IFIP Congress 68, August 1968.