Theorems on redundancies in distributed computing

Sahasrajit Sarmasarkar, Harish Pillai
Department of Electrical Engineering
Indian Institute of Technology, Bombay

Abstract

We consider the problem of redundancies in distributed computing where a master server wishes to compute some tasks and is provided a few child servers to compute. We consider a noisy environment where some child servers may fail to communicate their results to the master. We attempt to distribute tasks to the servers so that master is able to get the results for most of the tasks even if a few servers fail to communicate. We formulate some conditions on the distribution such that the number of tasks received is the maximum and also show that constructions using "Balanced Incomplete Block Design" [1] attains this optimality.

I. Preliminaries

We consider a setup where a master has n jobs to compute and has c servers to do the computations. We further assume a noisy scenario where only a fraction of these servers are able to communicate with the master. Therefore, we need to introduce redundancies in the setup by assigning each job to multiple servers. In our study, we examine the expectation and variance on the number of distinct completed jobs that the master receives from the servers that were able to communicate, and in particular, we study assignment schemes that achieve the desired variances on the number of received distinct jobs.

II. Notations and symbols

Given a set of n jobs and c servers, we would like to study various assignments of jobs to different servers by the master server. More formally, let us denote the n jobs by $\mathcal{A} = \{a_1, \dots, a_n\}$ and c servers by $\mathcal{S} = \{s_1, s_2, \dots, s_c\}$. Any assignment ($D$) of jobs in $\mathcal{A}$ to servers in $\mathcal{S}$, can be equivalently represented by a bipartite graph $\mathcal{G}_D$ where the nodes denote the jobs and the servers while edges exist between nodes denoting job a_i and server s_j if job a_i is assigned to server s_j. Alternately, for a job assignment D, we can define an assignment matrix $A_D \in \{0,1\}^{n \times c}$ as given below.

Definition 1. *(Construction of A_D): Given an assignment of jobs in $\mathcal{A}$ to servers in $\mathcal{S}$, we define matrix $A_D \in \{0,1\}^{n \times c}$ as follows. :*

$$\begin{aligned} A_D[i,j] &= 1 \text{ if job } a_i \text{ is assigned to server } s_j \\ &= 0 \text{ otherwise} \end{aligned} \tag{1}$$

Observe that the matrix A_D represents the adjacency graph matrix for the bipartite graph $\mathcal{G}_D$.

We now specifically study balanced assignment schemes where each server is assigned the same number of jobs and each job is assigned to the same number of servers. More formally, we define it as follows.

Definition 2. *(Balanced (n,k,r,c) assignment): Given a set of n jobs and c servers, we call an assignment scheme of jobs a balanced (n,k,r,c) assignment if the following conditions are satisfied.*

- *Each server is assigned precisely k distinct jobs to compute.*
- *Each job is assigned to precisely r distinct servers.*

Note that this assignment scheme ensures that $n \times r = k \times c$.

We can equivalently define it in terms of matrix A_D as follows.

Definition. *(Balanced (n, k, r, c) assignment in terms of A_D): Given a set of n jobs and c servers, we call the assignment scheme D of jobs to servers a balanced (n, k, r, c) assignment if each row of A_D sums up to r and each column sums up to k.*

Let us now describe an example of a balanced $(9, 3, 2, 6)$ assignment scheme.

Example 1. *We describe a balanced assignment scheme with 9 jobs $\{a_1, a_2, \ldots, a_9\}$ and 6 servers $\{s_1, s_2, \ldots, s_6\}$ in Table 1. Note that each job is assigned to precisely 2 servers and each server has exactly 3 jobs to compute. The assignment scheme is motivated from a cyclic assignment scheme*

Jobs	s_1	s_2	s_3	s_4	s_5	s_6
a_1	1	1				
a_2		1	1			
a_3			1	1		
a_4				1	1	
a_5					1	1
a_6	1					1
a_7	1	1				
a_8			1	1		
a_9					1	1

TABLE 1
ASSIGNMENT OF JOBS TO VARIOUS SERVERS IN A BALANCED $(9, 3, 2, 6)$ ASSIGNMENT SCHEME

III. THE MEAN AND THE VARIANCE

We consider the number of distinct jobs d received at the master when only a subset of x servers manage to communicate with the master. We consider any subset of $\mathcal{S}$ with cardinality x to be equally likely be the set of servers that communicates with the master. Note that with this definition, if $\hat{S} \subset \mathcal{S}$ (with $|\hat{S}| = x$) is the subset of servers that communicate with the master, then we can denote the number of distinct jobs received $d = | \cup_{j \in \hat{S}} \text{supp}(A_D[:, j])|$ where $\text{supp}(v)$ denotes the indices of the non-zero entries of the vector v.

Now, consider the uniform distribution over all subsets of servers of cardinality x which we denote by $\mathfrak{D}_{\mathcal{S},x}$ i.e. a sample from this distribution returns any subset of $\mathcal{S}$ of cardinality x with probability $\frac{1}{\binom{|\mathcal{S}|}{x}}$.

For a given assignment D of jobs to servers, we denote the expectation in the number of distinct completed jobs received by the master when any set of x servers is able to communicate with master uniformly at random by $\mathbb{E}_{D,x}[d]$ and the corresponding variance by $\sigma_{D,x}[d]$. The expectation and the variance on the number of distinct received jobs d may be written as

$$\mathbb{E}_{D,x}[d] = \mathbb{E}_{\hat{S} \sim \mathfrak{D}_{\mathcal{S},x}} \left[\left| \bigcup_{j \in \hat{S}} \text{supp}(A_D[:, j]) \right| \right] \text{ and } \sigma_{D,x}[d] = \sigma_{\hat{S} \sim \mathfrak{D}_{\mathcal{S},x}} \left[\left| \bigcup_{j \in \hat{S}} \text{supp}(A_D[:, j]) \right| \right] \quad (2)$$

Note that the randomness in this setup is only in the set of servers that can communicate with the master. The assignment scheme has no randomness associated with it.

Theorem 6 states that the expectation on the number of distinct jobs $\mathbb{E}_{D,x}[d]$ is the same for every balanced (n, k, r, c) assignment. This expectation is a function of n, k, r and c and is independent of the specific balanced assignment D we choose.

Theorem 1. *Consider any balanced (n, k, r, c) assignment D. The expectation of the number of distinct completed jobs d received by the master when any subset of servers $\mathcal{S}$ of cardinality x is able to communicate with the master with equal probability is the same for every assignment D amongst all balanced (n, k, r, c) assignments and is given by*

$$\mathbb{E}_{D,x}[d] = n.\left(1 - \frac{\binom{c-r}{x}}{\binom{c}{x}}\right) \tag{3}$$

Proof. We denote by $\mathfrak{n}^D_{i,\hat{S}}$ the number of servers in $\hat{S}$ to which job a_i is assigned under the assignment scheme D. Observe that $\mathfrak{n}^D_{i,\hat{S}}$ can take any value from 0 to r.

Now, the number of distinct jobs d received by the master when servers in a subset $\hat{S}$ (with $|\hat{S}| = x$) is able to communicate with the master is given by

$$d = \left|\bigcup_{j\in\hat{S}} \text{supp}(A_D[:,j])\right| = \left(k \times x - \sum_{i=1}^{n}(\mathfrak{n}^D_{i,\hat{S}} - 1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}\right) \tag{4}$$

Note that the term $\sum\limits_{i=1}^{n}(\mathfrak{n}^D_{i,\hat{S}} - 1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}$ excludes those jobs which have been received multiple times from various servers present in $\hat{S}$.

$$\begin{aligned}
\mathbb{E}_{D,x}[d] =& \frac{\sum\limits_{\substack{\hat{S}\subset\mathcal{S};\\ |\hat{S}|=x}} (k \times x - \sum\limits_{i=1}^{n}(\mathfrak{n}^D_{i,\hat{S}} - 1)\mathbb{1}_{\mathfrak{n}^D_{i,S}>1})}{\sum\limits_{\substack{\hat{S}\subset\mathcal{S};\\ |\hat{S}|=x}} 1} \\
=& k \times x - \frac{\sum\limits_{i=1}^{n}\sum\limits_{\substack{\hat{S}\subset\mathcal{S};\\ |\hat{S}|=x}} (\mathfrak{n}^D_{i,\hat{S}} - 1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}}{\binom{c}{x}} \\
=& k \times x - \frac{n\sum\limits_{\substack{\hat{S}\subset\mathcal{S};\\ |\hat{S}|=x}} (\mathfrak{n}^D_{i,\hat{S}} - 1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}}{\binom{c}{x}}
\end{aligned} \tag{5}$$

Observe that for every job a_i in a balanced (n, k, r, c) assignment, $\sum\limits_{\hat{S}\subset\mathcal{S},|\hat{S}|=x}(\mathfrak{n}^D_{i,\hat{S}} - 1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}$ is the same, i.e., this summation is independent of i. Now we show $\sum\limits_{\hat{S}\subset\mathcal{S},|\hat{S}|=x}(\mathfrak{n}^D_{i,\hat{S}} - 1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}$ is same for every balanced (n, k, r, c) distribution D for any specified x. We compute this sum by counting the number of subsets $\hat{S} \subset \mathcal{S}$ of cardinality x which additionally satisfies the constraint on $\mathfrak{n}^D_{i,\hat{S}} = t$ (i.e. job a_i is present in exactly t servers from $\hat{S}$) for every t from 2 to r (as these cases deal with the job a_i appearing more than once in the subset $\hat{S}$).

$$\sum_{\substack{\hat{S}\subset\mathcal{S};\\|\hat{S}|=x}}(\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}=\sum_{t=1}^{r-1}\sum_{\substack{\hat{S}\subset\mathcal{S};\\|\hat{S}|=x,\mathfrak{n}^D_{i,\hat{S}}=t+1}} t$$

$$=\sum_{t=1}^{r-1} t \sum_{\substack{\hat{S}\subset\mathcal{S};\\|\hat{S}|=x,\mathfrak{n}^D_{i,\hat{S}}=t+1}} 1$$

$$\overset{(a)}{=}\sum_{t=1}^{r-1} t\binom{r}{(t+1)}\binom{(c-r)}{(x-t-1)} \tag{6}$$

The last equality (a) comes from counting the number of subsets $\hat{S}\subset\mathcal{S}$ of cardinality x that contain precisely $t+1$ servers assigned the job a_i. Consider the following binomial expressions

$$ry(1+y)^{r-1}+1-(1+y)^r=\sum_{t=0}^{r-1} t\binom{r}{t+1}y^{t+1} \tag{7}$$

$$(1+y)^{c-r}=\sum_{u=0}^{c-r}\binom{c-r}{u}y^u \tag{8}$$

Multiplying equations (7) and (8), we can show that $\sum\limits_{t=1}^{r-1} t\binom{r}{(t+1)}\binom{(c-r)}{(x-t-1)}$ is precisely the coefficient of y^x in $ry(1+y)^{c-1}+(1+y)^{c-r}-(1+y)^c$. Thus,

$$\sum_{t=1}^{r-1} t\binom{r}{(t+1)}\binom{(c-r)}{(x-t-1)}=r\times\binom{c-1}{x-1}+\binom{c-r}{x}-\binom{c}{x} \tag{9}$$

Combining equations (5), (6) and (9), we get

$$\mathbb{E}_{D,x}[d]=k\times x-\frac{n\left(r\times\binom{c-1}{x-1}+\binom{c-r}{x}-\binom{c}{x}\right)}{\binom{c}{x}}=n\left(1-\frac{\binom{c-r}{x}}{\binom{c}{x}}\right)$$

■

A few comments are in order here. Note that for $x=1$, the expectation (as expected) is precisely k. Observe that for $x>c-r$, the expectation goes to n. In other words, if the number of servers that successfully communicates with the master is greater than $(c-r)$, then the master obtains at least one copy of every job $a_i\in\mathcal{A}$. This follows since every job is assigned to exactly r servers and therefore for any job to be missed out, the r servers to which that specific job was assigned, should fail to communicate with the master. Thus, if any job is missed out, then the number of servers that manage to communicate with the master x can at most be $c-r$.

We now calculate the variance for the number of distinct jobs d received at the master for any balanced (n,k,r,c) job assignment D. From the comments in the previous paragraph, it is clear that $\sigma_{D,1}[d]=0$ for the case $x=1$, since the master always receives precisely k distinct jobs, if only one server manages to communicate with the master. Similarly, $\sigma_{D,x}[d]=0$ for $x>c-r$, as the master would receive all the n jobs if more than $c-r$ servers communicate as we show in Corollary 1.

For calculating the variance on the number of distinct jobs d received by the master, observe

$$\sigma_{D,x}(d)=\sigma_{D,x}(k\times x-\sum_i(\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1})=\sigma_{D,x}\left(\sum_i(\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}\right)$$

The above follows since $\sigma(c-X)=\sigma(X)$ where c is a constant and X is a random variable. We now make use of the definition $\text{var}(X)=\mathbb{E}[X^2]-(\mathbb{E}[X])^2$. Therefore,

$$\begin{aligned}
\sigma_{D,x}(d) &= \sigma_{D,x}\left(\sum_i (\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}\right) \\
&= \frac{\sum\limits_{\substack{\hat{S}\subset\mathcal{S};\\|\hat{S}|=x}}\left(\sum\limits_{i=1}^{n}(\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}\right)^2}{\sum\limits_{\substack{\hat{S}\subset\mathcal{S};\\|\hat{S}|=x}}1} - \left(\frac{\sum\limits_{\hat{S}\subset\mathcal{S};|\hat{S}|=x}(\sum\limits_{i=1}^{n}(\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1})}{\sum\limits_{\substack{\hat{S}\subset\mathcal{S};\\|\hat{S}|=x}}1}\right)^2 \\
&\overset{(a)}{=} \frac{\sum\limits_{\substack{\hat{S}\subset\mathcal{S};\\|\hat{S}|=x}}(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}(\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}(\mathfrak{n}^D_{j,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{j,\hat{S}}>1})}{\sum\limits_{\substack{\hat{S}\subset\mathcal{S};\\|\hat{S}|=x}}1} - \left(\frac{\sum\limits_{i=1}^{n}\sum\limits_{\substack{\hat{S}\subset\mathcal{S};\\|\hat{S}|=x}}((\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1})}{\binom{c}{x}}\right)^2 \\
&\overset{(b)}{=} \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}\sum\limits_{\substack{\hat{S}\subset\mathcal{S};\\|\hat{S}|=x}}(\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}(\mathfrak{n}^D_{j,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{j,\hat{S}}>1}}{\binom{c}{x}} - \left(\frac{n\sum\limits_{t=1}^{r-1}t\binom{r}{(t+1)}\binom{(c-r)}{(x-t-1)}}{\binom{c}{x}}\right)^2 \\
&\overset{(c)}{=} \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}\sum\limits_{\substack{\hat{S}\subset\mathcal{S};\\|\hat{S}|=x}}(\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}(\mathfrak{n}^D_{j,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{j,\hat{S}}>1}}{\binom{c}{x}} - \left(\frac{n\binom{c-r}{x}}{\binom{c}{x}}+n\times\left(\frac{rx}{c}-1\right)\right)^2
\end{aligned} \tag{10}$$

In the above set of equations, (a) follows from the identity $(\sum\limits_i b_i)^2=\sum\limits_i\sum\limits_j b_ib_j$. The first term in (b) is obtained by interchanging the order of summations, whereas the second term comes from equation (6). Further, the second term in (c) follows using the equation (9) in the proof of Theorem 6.

Observe that the second term in the final expression in equation (10) depends only on n,r,c and x and is independent of the specific balanced (n,k,r,c) job assignment D. On the other hand, the first term in equation (10) depends on the particular assignment D. We now consider the numerator of the first term of equation (10) in more detail. We can break this expression into two parts, where one part is dependent on just one index i and the other part is dependent on two distinct indices i,j. Thus,

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{\hat{S}\subset\mathcal{S};\\|\hat{S}|=x}}(\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}(\mathfrak{n}^D_{j,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{j,\hat{S}}>1} \\
=&2\sum_{1\le i<j\le n}\sum_{\substack{\hat{S}\subset\mathcal{S};\\|\hat{S}|=x}}(\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}(\mathfrak{n}^D_{j,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{j,\hat{S}}>1}+\sum_{1\le i=j\le n}\sum_{\substack{\hat{S}\subset\mathcal{S};\\|\hat{S}|=x}}(\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}(\mathfrak{n}^D_{j,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{j,\hat{S}}>1}
\end{aligned} \tag{11}$$

In equation (11), the second term can be rewritten as $\sum\limits_{1\le i\le n}\sum\limits_{\substack{\hat{S}\subset\mathcal{S};\\|\hat{S}|=x}}((\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1})^2$. For every job a_i, this expression calculates $\sum\limits_{\substack{\hat{S}\subset\mathcal{S};\\|\hat{S}|=x}}((\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1})^2$, which is independent of the choice of the job a_i in any balanced (n,k,r,c) assignment D. In fact, this second term of equation (11) is independent of the choice of D and it depends only on the values of c, r and x. We can compute this sum by counting the number of subsets $\hat{S} \subset \mathcal{S}$ of cardinality x that additionally satisfy the constraint $\mathfrak{n}^D_{i,\hat{S}} = t$ (i.e. job a_i is present in exactly t servers from $\hat{S}$) for every t from 2 to r. Thus

$$\sum_{1\le i\le n} \sum_{\substack{\hat{S}\subset\mathcal{S};\\ |\hat{S}|=x}} ((\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1})^2 = n\sum_{t=1}^{r-1} t^2 \binom{r}{(t+1)}\binom{(c-r)}{(x-t-1)} \tag{12}$$

Note that the number of subsets $\hat{S} \subset \mathcal{S}$ of cardinality x such that a particular job a_i appears $t+1$ times in $\hat{S}$ is given by $\binom{r}{(t+1)}\binom{(c-r)}{(x-t-1)}$. As $\mathfrak{n}^D_{i,\hat{S}} = t+1$ for this particular $\hat{S}$, therefore $((\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1})^2 = t^2$. This explains the final expression in equation (12). A closed form expression for the sum $\sum_{t=1}^{r-1} t^2\binom{r}{(t+1)}\binom{(c-r)}{(x-t-1)}$ can be obtained by considering the following binomial expressions

$$r(r-1)y^2(1+y)^{r-2} - ry(1+y)^{r-1} - 1 + (1+y)^r = \sum_{t=1}^{r-1} t^2\binom{r}{t+1} y^{t+1} \tag{13}$$

$$(1+y)^{c-r} = \sum_{v=0}^{c-r}\binom{c-r}{v} y^v \tag{14}$$

Multiplying equations (13) and (14), one obtains $\sum_{t=1}^{r-1} t^2\binom{r}{t+1}\binom{c-r}{x-t-1}$ to be the coefficient of y^x in $r(r-1)y^2(1+y)^{c-2} - ry(1+y)^{c-1} - (1+y)^{c-r} + (1+y)^c$, thus,

$$\sum_{t=1}^{r-1} t^2\binom{r}{t+1}\binom{c-r}{x-t-1} = r(r-1)\binom{c-2}{x-2} - r\binom{c-1}{x-1} - \binom{c-r}{x} + \binom{c}{x} \tag{15}$$

Finally, we analyse the first term in equation (11), viz., $\sum_{1\le i<j\le n}\sum_{\substack{\hat{S}\subset\mathcal{S};\\ |\hat{S}|=x}} (\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}(\mathfrak{n}^D_{j,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{j,\hat{S}}>1}$.

If a specific pair of jobs a_i, a_j appear $(\alpha+1)$ and $(\beta+1)$ times respectively in some subset $\hat{S} \subset \mathcal{S}$ of cardinality x, then such a pair of jobs contribute $\alpha\beta$ towards this expression that we are analysing. One needs to add up such contributions from every distinct pair of jobs (a_i, a_j) and every subset $\hat{S} \subset \mathcal{S}$ of cardinality x to get the final value of this expression. The strategy that we adopt to compute this sum is as follows : we find $\sum_{\substack{\hat{S}\subset\mathcal{S};\\ |\hat{S}|=x}} (\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}(\mathfrak{n}^D_{j,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{j,\hat{S}}>1}$ for any given pair of distinct jobs (a_i, a_j). Observe that this expression depends on how the pair of jobs (a_i, a_j) are distributed amongst the c servers, which in turn depends on the particular balanced (n,k,r,c) job assignment D that is under consideration. Now, given a particular pair of jobs (a_i, a_j), how they are farmed to the servers can essentially differ only in the number of servers that are assigned both the jobs a_i, a_j simultaneously. The number of servers that are simultaneously assigned both the jobs (a_i, a_j) can range from 0 to r. If a pair of jobs (a_i, a_j) are assigned together to precisely m servers (with $0 \le m \le r$), then the sum $\sum_{\substack{\hat{S}\subset\mathcal{S};\\ |\hat{S}|=x}} (\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}(\mathfrak{n}^D_{j,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{j,\hat{S}}>1}$ calculated for this particular pair of jobs is precisely equal to the corresponding sum for every other pair of jobs that are assigned together to precisely m servers. We use the notation

$$g(m,x) = \sum_{\substack{\hat{S}\subset\mathcal{S};\\ |\hat{S}|=x}} (\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}(\mathfrak{n}^D_{j,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{j,\hat{S}}>1} \tag{16}$$

to indicate this particular sum. We now show that the values of $g(m,x)$ depends only on c,r,m and x. We give the expression for $g(0,x)$ in Lemma 1 and give a recursion to compute $g(m,x)$ in Lemma 2.

Lemma 1. *For a balanced* (n,k,r,c) *assignment, the value of* $g(0,x)$ *is given by*

$$g(0,x) = r^2\binom{c-2}{x-2} - 2r\binom{c-1}{x-1} + \binom{c}{x} - 2\binom{c-r}{x} + 2r\binom{c-r-1}{x-1} + \binom{c-2r}{x} \tag{17}$$

Proof. Consider a pair of jobs (a_i,a_j) such that no server has been assigned both a_i and a_j together. Therefore there are precisely r servers that have been assigned a_i and not a_j. Another r servers that are assigned a_j but not a_i and $c-2r$ servers that are assigned neither a_i nor a_j. Then

$$g(0,x) = \sum_{\substack{\hat{S}\subset\mathcal{S};\\ |\hat{S}|=x}} (\mathfrak{n}^D_{i,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}(\mathfrak{n}^D_{j,\hat{S}}-1)\mathbb{1}_{\mathfrak{n}^D_{j,\hat{S}}>1} = \sum_{t=2}^{r}\sum_{u=2}^{r}(t-1)(u-1)\binom{r}{t}\binom{r}{u}\binom{c-2r}{x-t-u} \tag{18}$$

Clearly any subset of servers $\hat{S}\subset\mathcal{S}$ of cardinality x that has at most only one instance of the job a_i assigned amongst its members does not contribute to the sum. Ditto for a_j. Therefore, one needs to consider only those subsets $\hat{S}$ of servers that contain at least two servers that are assigned a_i and at least two servers that are assigned a_j. In the final expression of equation (18), $\binom{r}{t}\binom{r}{u}\binom{c-2r}{x-t-u}$ counts the number of subsets of servers $\hat{S}$ of cardinality x that contain t servers assigned a_i, u servers assigned a_j and $x-t-u$ servers that have been assigned neither. The summation limits ensure that there are at least 2 servers assigned a_i and at least 2 servers assigned a_j. The expression $(t-1)(u-1)$ is the contribution of each subset $\hat{S}$ that contains t copies of a_i and u copies of a_j assigned to its members. A closed form solution of the expression for $g(0,x)$ can be obtained by considering

$$ry(1+y)^{r-1} + 1 - (1+y)^r = \sum_{t=1}^{r}(t-1)\binom{r}{t}y^t \tag{19}$$

$$ry(1+y)^{r-1} + 1 - (1+y)^r = \sum_{u=1}^{r}(u-1)\binom{r}{u}y^u \tag{20}$$

$$(1+y)^{c-2r} = \sum_{v=0}^{c-2r}\binom{c-2r}{v}y^v \tag{21}$$

Multiplying these three expressions (19), (20) and (21), we get $\sum_{t=2}^{r}\sum_{u=2}^{r}(t-1)(u-1)\binom{c-2r}{x-t-u}\binom{r}{t}\binom{r}{u}$ to be the coefficient of y^x in $\left(ry(1+y)^{r-1}+1-(1+y)^r\right)^2(1+y)^{c-2r}$.

$$\begin{aligned}&\sum_{t=2}^{r}\sum_{u=2}^{r}(t-1)(u-1)\binom{c-2r}{x-t-u}\binom{r}{t}\binom{r}{u}\\ =& r^2\binom{c-2}{x-2} - 2r\binom{c-1}{x-1} + \binom{c}{x} - 2\binom{c-r}{x} + 2r\binom{c-r-1}{x-1} + \binom{c-2r}{x}\end{aligned} \tag{22}$$

■

Lemma 2. *For a balanced* (n,k,r,c) *assignment, the values for* $g(m,x)$ *are related in the following fashion*

$$g(m+1,x) - g(m,x) = \binom{c-2}{x-1} - 2\binom{c-r-1}{x-1} + \binom{c-2r+m}{x-1} \tag{23}$$

Proof. Let us consider a pair of jobs (a_i,a_j) that have been assigned together to precisely m servers. Without loss of generality, let $s_1,s_2,\cdots,s_m$ be the servers that are assigned both the jobs a_i,a_j. Let

$s_{m+1}, s_{m+2}, \cdots, s_r$ be the servers that have been assigned a_i but not a_j. Assume servers $s_{r+1}, s_{r+2}, \cdots, s_{2r-m}$ are the servers assigned a_j but not a_i. The last $c-2r+m$ servers $s_{2r-m+1}, s_{2r-m+2}, \cdots, s_c$ are the ones that have not been assigned a_i or a_j.

Let another pair of jobs (a_{i_1}, a_{j_1}) be such that they have been assigned together to precisely $m+1$ servers. We now consider a bijective map $f : \mathcal{S} \to \mathcal{S}$ described in the following fashion. Let $f(s_\ell)$ for $1 \le \ell \le m+1$ be servers that have been assigned both the jobs a_{i_1} and a_{i_2}. Let $f(s_\ell)$ for $m+2 \le \ell \le r$ be servers that have been assigned a_{i_1} but not a_{j_1}. Further let $f(s_\ell)$ for $r+2 \le \ell \le 2r-m$ be servers that have been assigned a_{j_1} but not a_{i_1}. The rest of $f(s_\ell)$ have not been assigned a_{i_1} or a_{j_1}. Thus there are two special servers, namely s_{m+1} (which does job a_i) and s_{r+1} (which does job a_j), and whose images $f(s_{m+1})$ (which does both the jobs a_{i_1} and a_{j_1}) and $f(s_{r+1})$ (which does neither a_{i_1} nor a_{j_1}) that we shall pay special attention.

For any $\hat{S} \subset \mathcal{S}$ of cardinality x, let us compare its contribution to the sum $g(m, x)$ with the contribution of $f(\hat{S})$ towards $g(m+1, x)$. Clearly, if $\hat{S} \subset \mathcal{S} \setminus \{s_{m+1}, s_{r+1}\}$, then the contribution of $\hat{S}$ towards $g(m, x)$ is exactly the same as the contribution of $f(\hat{S})$ to $g(m+1, x)$. Similarly, if $s_{m+1}, s_{r+1} \in \hat{S}$, then contribution of $\hat{S}$ towards $g(m, x)$ and that of $f(\hat{S})$ is exactly the same. Therefore it suffices to only consider those subsets $\hat{S}$ of cardinality x that contain exactly one of the two special servers $\{s_{m+1}, s_{r+1}\}$ to evaluate the difference $g(m+1, x) - g(m, x)$. Hence we look at subsets $\hat{S}$ that are formed by taking either s_{m+1} or s_{r+1} along with $\bar{S} \subset \mathcal{S} \setminus \{s_{m+1}, s_{r+1}\}$ of cardinality $x-1$.

Let $\bar{S} \subset \mathcal{S} \setminus \{s_{m+1}, s_{r+1}\}$ of cardinality $x-1$ contain $\alpha > 0$ instances of job a_i and $\beta > 0$ instances of job a_j assigned to its servers. Then $\bar{S} \cup \{s_{m+1}\}$ contributes $\alpha(\beta-1)$ towards $g(m, x)$, whereas $\bar{S} \cup \{s_{r+1}\}$ contributes $(\alpha-1)\beta$ towards $g(m, x)$. At the same time, $f(\bar{S}) \cup \{f(s_{m+1})\}$ contributes $\alpha\beta$ towards $g(m+1, x)$, whereas $f(\bar{S}) \cup \{f(s_{r+1})\}$ contributes $(\alpha-1)(\beta-1)$ towards $g(m+1, x)$. Thus, one can evaluate the contribution of $\bar{S}$ towards the difference $g(m+1, x) - g(m, x)$ to be $\alpha\beta + (\alpha-1)(\beta-1) - \alpha(\beta-1) - (\alpha-1)\beta = 1$. So every subset $\bar{S} \subset \mathcal{S} \setminus \{s_{m+1}, s_{r+1}\}$ of cardinality $x-1$, whose servers have at least one instance each of jobs a_i and a_j assigned to them, contributes a net change of 1 towards the difference $g(m+1, x) - g(m, x)$. One needs to just count the number of subsets $\bar{S}$ of cardinality $x-1$ that satisfy these conditions to find $g(m+1, x) - g(m, x)$.

Total number of subsets of cardinality $x-1$ of the set $\mathcal{S} \setminus \{s_{m+1}, s_{r+1}\}$ of cardinality $c-2$ is given by $\binom{c-2}{x-1}$. If the subset $\bar{S}$ is one of the $\binom{c-r-1}{x-1}$ subsets chosen from the servers $\{s_{r+2}, s_{r+2}, \cdots s_c\}$, then the job a_i is not assigned to any of its servers. Similarly, if $\bar{S}$ is one of the $\binom{c-r-1}{x-1}$ chosen from the servers $\{s_{m+2}, s_{m+3}, \cdots, s_r\} \cup \{s_{2r-m+1}, s_{2r-m+2}, \cdots, s_c\}$ does not have any instance of the job a_j assigned to its servers. As these subsets $\bar{S}$ do not contribute to the difference $g(m=1, x) - g(m, x)$, their numbers have to be subtracted from $\binom{c-2}{x-1}$. In the process, $\bar{S} \subset \{s_{2r-m+1}, s_{2r-m+2}, \cdots, s_c\}$ have been subtracted twice and therefore $\binom{c-2r+m}{x-1}$ needs to added back, thereby giving

$$g(m+1, x) - g(m, x) = \binom{c-2}{x-1} - 2\binom{c-r-1}{x-1} + \binom{c-2r+m}{x-1}$$

■

Note that all the expressions for $g(m, x)$ depends on the values of c, r, m and x and is therefore independent of which balanced (n, k, r, c) assignment D we choose.

We now define $\mathfrak{m}^D(m)$ as the number of distinct pairs of jobs (a_i, a_j) with $1 \le i < j \le n$ that are assigned together to precisely m servers in the balanced (n, k, r, c) assignment D. One can formally define this number for a specific balanced (n, k, r, c) assignment D using the assignment matrix A_D as

$$\mathfrak{m}^D(m) = \sum_{\substack{(i_1, i_2) \\ 1 \le i_1 < i_2 \le n}} \mathbb{1}_{\sum_{j=1}^{c} A_D[i_1, j] A_D[i_2, j] = m} \tag{24}$$

Given a balanced (n, k, r, c) assignment D, the numbers $\mathfrak{m}^D(m)$ have some additional properties :

$$\sum_{m=0}^{r} \mathfrak{m}^D(m) = \binom{n}{2} \tag{25}$$

$$\sum_{m=0}^{r} m\mathfrak{m}^D(m) = c\binom{k}{2} \tag{26}$$

Equation (25) follows from the fact that there are a total of n jobs and thus the total number of such pairs is given by $\binom{n}{2}$. Equation (26) follows from the fact that the number of pairs of jobs that are assigned together to a fixed server s_i is given by $\binom{k}{2}$. Summing over all such servers in $\mathcal{S}$ gives us the RHS in (26). Note that we count each pair of jobs as many times as they appear together in a server and thus we say $\sum_{m=0}^{r} m\mathfrak{m}^D(m) = c\binom{k}{2}$.

Observe that the first term of equation (11) that we are evaluating can now be written as

$$\sum_{1\leq i<j\leq n} \sum_{\substack{\hat{S}\subset\mathcal{S};\\ |\hat{S}|=x}} (\mathfrak{n}^D_{i,\hat{S}} - 1)\mathbb{1}_{\mathfrak{n}^D_{i,\hat{S}}>1}(\mathfrak{n}^D_{j,\hat{S}} - 1)\mathbb{1}_{\mathfrak{n}^D_{j,\hat{S}}>1} = \sum_{m=0}^{r} \mathfrak{m}^D(m)g(m,x) \tag{27}$$

We therefore now conclude the following result regarding evaluation of variance $\sigma_{D,x}(d)$:

Theorem 2. *Consider any assignment D amongst balanced (n,k,r,c) assignments. The variance on the number of distinct jobs d $(\sigma_{D,x}(d))$ received at the master when any subset of servers $\mathcal{S}$ of cardinality $x>1$ is able to communicate to the master with equal probability is stated as an equation below.*

$$\sigma_{D,x}(d) = \frac{2\sum_{m=0}^{r} \mathfrak{m}^D(m)g(m,x) + T_2(n,r,c,x)}{\binom{c}{x}} - (T_1(n,r,c,x))^2 \tag{28}$$

where $T_1(n,r,c,x) = \frac{n\binom{c-r}{x}}{\binom{c}{x}} + n\times(\frac{rx}{c} - 1)$

$$T_2(n,r,c,x) = n\sum_{t=1}^{r-1} t^2\binom{r}{(t+1)}\binom{(c-r)}{(x-t-1)} \tag{29}$$

$$= n\left(r(r-1)\binom{c-2}{x-2} - r\binom{c-1}{x-1} - \binom{c-r}{x} + \binom{c}{x}\right) \tag{30}$$

Thus, we have shown that while the mean of the number of received jobs d is the same for all balanced (n,k,r,c) assignments, the variance of d is dependent on the frequency of job pairs assigned to the servers.

IV. Results on the least variance on the number of distinct jobs d

With this setup, we now study assignment schemes on when the least variance on the number of distinct jobs d is attained. We specifically propose pairwise balanced designs discussed in Definition 4 and 5 in Theorem 3 and 4 and show that it attains the least variance amongst the class of balanced assignments. We now compare this definition to balanced incomplete block designs (BIBD) [1], [2] and show that it is a generalization of the same.

Defintion 3. *(BIBD (v,b,r,k,λ) scheme as in [2]) - A balanced incomplete block design (BIBD) is a pair (V,B) where V is a v-set and B is a collection of b k-sized subsets of V (blocks) such that each element of V is contained in exactly r blocks and any 2-subset of V is contained in exactly λ blocks.*

Note that this set up can be directly mapped to our setup where V denotes the set of all jobs and each block in B denotes a server and each server(block) in B is precisely assigned those jobs corresponding to the set of elements of V in that block. Also, observe that in BIBD design, we require every set of jobs to occur together in the same number of servers but our pairwise balanced schemes allow a tolerance of upto 1 as we can see in Definition 4.

Definition 4. *(Pairwise balanced job (n,k,r,c) assignment): Given a balanced (n,k,r,c) assignment, we call it pairwise balanced job (n,k,r,c) assignment if for every pair of distinct jobs a_i and a_j in $\mathcal{A}$, the number of servers which are assigned both a_i and a_j is exactly l or $l+1$ for some integer l.*

Using $\mathfrak{m}^D(.)$ defined in (24), we can define pairwise balanced job assignments below. W.L.O.G, we may assume in pairwise balanced job (n,k,r,c) assignment, there always exist a pair of jobs occurring together in l server and there may or may not be pairs of jobs that occur together in $l+1$ servers.

Definition (Definition 4 in terms of $\mathfrak{m}^D(.)$)**.** *(Pairwise balanced job (n,k,r,c) assignment): Given a balanced (n,k,r,c) assignment, we call it server pairwise -balanced (n,k,r,c) if $\mathfrak{m}_D(m)$ is non-zero for at most 2 consecutive values of m (call it l and $l+1$).*

Thus, we can see that BIBD would also fall into the category of pairwise balanced job assignment schemes. The existence of BIBD is still not fully understood and some sufficient conditions on n,k,r and c under which BIBD exists are also discussed in [1], [2]. The famous Bruck-Ryser-Chowla theorem gives some necessary condition on n,k,c,r for the existence of BIBDs.

The following corollary proves a result on l for any pairwise balanced job (n,k,r,c) assignment.

Lemma 3. *For any pairwise balanced job (n,k,r,c) assignment, we have $l = \left\lfloor \frac{c.\binom{k}{2}}{\binom{n}{2}} \right\rfloor$*

Proof. For any pairwise balanced job (n,k,r,c) assignment D, we must have $\mathfrak{m}^D(m)$ to be zero for all $m \neq l, l+1$. Thus, we have $\mathfrak{m}^D(l+1) = \binom{n}{2} - \mathfrak{m}^D(l)$ using eqaution (25) and similarly, we have $l(\mathfrak{m}^D(l)) + (l+1)(\binom{n}{2} - \mathfrak{m}^D(l)) = c\binom{k}{2}$ using equation (26). We thus have $l\binom{n}{2} \leq c\binom{k}{2}$ and $(l+1)\binom{n}{2} > c\binom{k}{2}$ (as $\mathfrak{m}^D(l) > 0$) and thus $l = \left\lfloor \frac{c.\binom{k}{2}}{\binom{n}{2}} \right\rfloor$

■

Definition 5. *(Pairwise balanced server (n,k,r,c) assignment): Given a balanced (n,k,r,c) assignment, we call it server pairwise -balanced (n,k,r,c) if for every pair of distinct servers s_i and s_j in $\mathcal{S}$, the number of jobs assigned to both s_i and s_j is exactly l or $l+1$ for some integer l.*

Lemma 4. *For any pairwise balanced server (n,k,r,c), we must have $l = \left\lfloor \frac{n.\binom{r}{2}}{\binom{c}{2}} \right\rfloor$*

This lemma can be proved in a very similar way as that of Lemma 3.

With this, we could alternately redefine pairwise balanced server (n,k,r,c) schemes as follows in terms of matrix A_D as follows.

Definition (Definition 4 in terms of A_D)**.** *(Pairwise balanced server (n,k,r,c) assignment): Given a balanced (n,k,r,c) assignment, we call it server pairwise -balanced (n,k,r,c) if for every pair of distinct indices i and j, $|supp(A_D[:,i]) \cap supp(A_D[:,j])| = l$ or $l+1$ for some integer l.*

Definition (Definition 5 in terms of A_D)**.** *(Pairwise balanced job (n,k,r,c) assignment): Given a balanced (n,k,r,c) assignment, we call it job pairwise -balanced (n,k,r,c) if for every pair of distinct indices i and j, $|supp(A_D[i,:]) \cap supp(A_D[j,:])| = l$ or $l+1$ for some integer l.*

We now describe a pairwise balanced job and server scheme in Example 2 below. Note that the balanced design is both pairwise job and server balanced in this design. Further, in example 3, we present an assignment scheme which is pairwise server balanced but not pairwise job balanced.

Example 2. *We now describe a pairwise balanced server and job $(9,3,3,9)$ assignment scheme in Table 2. Note that in this scheme, every pair of jobs appear together in exactly 1 or 0 servers, thus it is a pairwise balanced job assignment scheme. Similarly, we can also show that it is also a pairwise balanced server scheme too as every pair of servers has either one job or zero jobs common.*

Jobs	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9
a_1	1	1	1						
a_2				1	1	1			
a_3							1	1	1
a_4	1			1			1		
a_5		1			1			1	
a_6			1			1			1
a_7	1				1				1
a_8		1				1	1		
a_9			1	1				1	

TABLE 2
ASSIGNMENT OF JOBS TO VARIOUS SERVERS IN A PAIRWISE BALANCED $(9, 3, 2, 6)$ ASSIGNMENT SCHEME

Example 3. *We now describe a pairwise balanced server $(14, 6, 3, 7)$ assignment but not a pairwise balanced job scheme in Table 3. Note that in this scheme, every pair of servers have exactly 2 jobs in common. However, some pairs of jobs appear together in 1 server like (a_1, a_2) and some pairs of jobs appear together in 3 servers like pairs of jobs (a_1, a_8) and thus it is not pairwise job balanced.*

Jobs	s_1	s_2	s_3	s_4	s_5	s_6	s_7
a_1		1		1		1	
a_2	1			1	1		
a_3			1	1			1
a_4	1	1	1				
a_5		1			1		1
a_6	1					1	1
a_7			1		1	1	
a_8		1		1		1	
a_9	1			1	1		
a_{10}			1	1			1
a_{11}	1	1	1				
a_{12}		1			1		1
a_{13}	1					1	1
a_{14}			1		1	1	

TABLE 3
ASSIGNMENT OF JOBS TO VARIOUS SERVERS IN A PAIRWISE SERVER BALANCED $(14, 6, 3, 7)$ ASSIGNMENT SCHEME

Note that the following two theorems say that the variance on the number of distinct d received at the master is the least for the pairwise balanced designs discussed in Definition 4 and 5. However such designs may not always exist and it is not entirely known when such designs exist for every tuple of (n, k, r, c) satisfying $n \times r = k \times c$. This is closely related to balanced incomplete block designs (BIBD) [1] whose existence is not fully and clearly understood. However, a key difference for the pairwise balanced designs we consider is that it is a generalization of BIBD in which they restrict the number of jobs common to any pair of servers to be the same for every such pair of servers.

Recall that for the following theorems, we study the setup described in Section III where any set of x servers could be the only set of servers that could communicate with the master. Recall that we further assume that every set of x servers is equally likely to be the only set of servers that could communicate with the master.

Theorem 3. *The least variance of the number of distinct jobs d received at the master for any $x \in [2, c - 2r + 1]$ is attained uniquely by the pairwise balanced job (n, k, r, c) assignment schemes (if it exists) amongst all balanced (n, k, r, c) assignments.*

However, for $x > c-2r+1$, the result on least variance still holds but the uniqueness is not guaranteed.

We prove this theorem in Section VIII

Also under a special constraint $n = c$, the above theorem can also be stated as follows and we can say that the same result holds true for pairwise server balanced (n, k, r, c) scheme too.

Theorem 4. *The least variance of the number of distinct jobs received at the master for any $x \in [1, c-2r+1]$ is attained uniquely by both the pairwise balanced server (n, k, k, n) assignment schemes and pairwise balanced job (n, k, k, n) assignment schemes (if it exists) amongst all balanced (n, k, k, n) assignments. However, for $x > c-2r+1$, the result on least variance still holds but the uniqueness is not guaranteed.*

We prove this theorem in Section X.

V. Results on the largest variance on the number of distinct jobs d

With this setup, we now study assignment schemes on when the largest variance in the number of distinct jobs d is attained. We specifically define pairwise heavy imbalanced job (n, k, r, c) assignment discussed in Definition 6 and show that in Theorem 5 that these schemes attain the largest variance amongst the class of balanced assignment schemes.

Definition 6. *(Pairwise heavy imbalanced job (n, k, r, c) assignment): Given a balanced (n, k, r, c) assignment, we call it pairwise imbalanced job (n, k, r, c) assignment if for every pair of distinct jobs a_i and a_j in $\mathcal{A}$, the number of servers which are assigned both a_i and a_j is exactly 0 or r.*

Definition (Definition 6 in terms of A_D)**.** *(Pairwise heavy imbalanced job (n, k, r, c) assignment): Given a balanced (n, k, r, c) assignment, we call it job pairwise heavy-imbalanced job (n, k, r, c) assignment if for every pair of distinct indices i and j, $|supp(A_D[i, :]) \cap supp(A_D[j, :])| = 0$ or r.*

Lemma 5. *A pairwise heavy imbalanced job (n, k, r, c) assignment can exist only when either n or $k-1$ is even.*

Proof. Observe from (26) that $\sum_{m=0}^{r} \mathfrak{m}^D(m) = c\binom{k}{2}$. Suppose neither n nor $k-1$ is even, thus implying that $n(k-1)$ is not divisible by 2. However, for a pairwise heavy imbalanced job (n, k, r, c) assignment scheme D, $\mathfrak{m}^D(m)$ is non-zero only for $m \neq 0, r$ and thus, $\mathfrak{m}^D(r) = \frac{c\binom{k}{2}}{r} = \frac{n(k-1)}{2}$. Thus, $n(k-1)$ being odd would imply a fractional value for $\mathfrak{m}^D(r)$ which is clearly not possible. ■

Example 4. *We now describe a pairwise heavy imbalanced job $(9, 3, 2, 6)$ assignment scheme in Table 4. Note that in this scheme, every pair of jobs appear together in exactly 2 or 0 servers, thus it is a pairwise heavy imbalanced job assignment scheme.*

Jobs	s_1	s_2	s_3	s_4	s_5	s_6
a_1	1	1				
a_2	1	1				
a_3	1	1				
a_4			1	1		
a_5			1	1		
a_6			1	1		
a_7					1	1
a_8					1	1
a_9					1	1

TABLE 4
Assignment of jobs to various servers in a pairwise heavy imbalanced $(9, 3, 2, 6)$ assignment scheme

Recall that for the following theorems, we study the setup described in Section III where any set of x servers could be the only set of servers that could communicate with the master. Recall that we further assume that every set of x servers is equally likely to be the only set of servers that could communicate with the master.

Theorem 5. *The largest variance of the number of distinct jobs d received at the master for any $x \in [2, c-2r+1]$ is attained uniquely by the pairwise balanced job (n,k,r,c) assignment schemes (if it exists) amongst all balanced (n,k,r,c) assignments.*

However, for $x > c-2r+1$, the result on largest variance still holds but the uniqueness is not guaranteed.

We prove this theorem in Section IX

VI. Generalization where the number of servers that return (x) is random

We now discuss an alternate scenario where each server is independently and equally likely to communicate with the master with probability p. Note that under this setup, the distribution of the number of servers x that could communicate is given by $\text{Bin}(c,p)$. Also, observe that conditioned on x, every subset of x sized servers is equally likely to communicate with the master. Under this setup, we can now state our results on mean and variance on the number of distinct jobs on variance below.

Theorem 6. *Consider any balanced (n,k,r,c) assignment D. We now consider a setup where each server is independently and equally likely to communicate with the master with probability p. The expectation of the number of distinct completed jobs d received is the same for every assignment D amongst all balanced (n,k,r,c) assignments and is given by*

$$\mathbb{E}_D[d] = n - n(1-p)^r \tag{31}$$

and the variance is given by

$$\sigma_D(d) = \sum_{x=0}^{c} \sigma_{D,x}(d) \binom{c}{x} p^x (1-p)^{c-x} \tag{32}$$

where $\sigma_{D,x}(d)$ is given by the expression in Equation (28).

Proof. Observe that under this setup, the number of servers that could communicate with the master x can be given by a binomial distribution $\text{Bin}(c,p)$. Also observe that under this setup conditioned on x, any set of x servers is equally likely to communicate with the master.

We can thus say that

$$\mathbb{E}_D[d] = \mathbb{E}_{x\sim\text{Bin}(c,p)} \mathbb{E}_{D,x}[d] \tag{33}$$

$$\stackrel{(a)}{=} \sum_{x=0}^{c} n \left(1 - \frac{\binom{c-r}{x}}{\binom{c}{x}}\right) \binom{c}{x} p^x (1-p)^{c-x} \tag{34}$$

$$= n - n(1-p)^r \sum_{x=0}^{c} \binom{c-r}{x} p^x (1-p)^{c-r-x} = n - n(1-p)^r \tag{35}$$

Note (a) follows from the expression of mean in Theorem 6. Using a very similar technique, we can prove a result of $\sigma_D(d)$ as well. ∎

We can actually generalize some of our results in Theorem 3 and 4 for a more generalized setup where the number of servers that return is not unique but is sampled from some distribution $\mathcal{P}$. However, we ensure that the subset S_1 is the set of servers that could communicate is equally likely as the subset of servers S_2 that could communicate if $|S_1| = |S_2|$. Formally, we study the setup where x is sampled from a distribution $\mathcal{P}$ and conditioned on x, any subset of x servers is equally likely to be the set of servers that could communicate with the master.

This precisely captures the case where every server is independently able to communicate to the master with probability p, in which case $\mathcal{P}$ would be given by $\text{Bin}(c,p)$

Theorem 7. *Let us consider $x \sim \mathcal{P}$. Conditioned on x, we study the setup where any set of x servers is equally likely to communicate with the master. Then both the pairwise balanced job (n,k,r,c) assignment schemes and pairwise balanced server (n,k,r,c) assignment schemes (if exists) attain the least variance on the number of distinct jobs received at master amongst all balanced (n,k,r,c) assignment schemes.*

Proof. Let us denote the number of distinct jobs when any set of x servers return uniformly at random by d. However, in our problem x itself might be sampled from a distribution $\mathcal{P}$. Let us denote the variance in this set-up under this assignment of jobs to servers (say D) by $\sigma_{D,x\sim\mathcal{P}}(d)$.

Now using law of variances, we can say that

$$\sigma_{D,x\sim\mathcal{P}}(d) = \mathbb{E}_{x\sim\mathcal{P}}[\sigma_{D,x}(d)] + \sigma_{x\sim\mathcal{P}}[\mathbb{E}_{D,x}(d)]$$

. Now consider assignments D and D_1 such that assignment D is a pairwise balanced job (n,k,r,c) assignment scheme and assignment D_1 could be any balanced (n,k,r,c) assignment scheme.

However, we know from Thoerem 4 that $\sigma_{D,x}(d) \leq \sigma_{D_1,x}(d)$ for every x and assignments D, D_1 such that assignment D is a pairwise balanced job (n,k,r,c) assignment scheme and assignment D_1 could be any balanced (n,k,r,c) assignment scheme.

We also know that $\mathbb{E}_{D,x}(d) = \mathbb{E}_{D_1,x}(d)$ from Thoerem 6. Combining the two properties, we get that $\sigma_{D,x\sim\mathcal{P}}(d) \leq \sigma_{D',x\sim\mathcal{P}}(d)$ thus proving the theorem. ∎

Using a very similar approach, we can also prove a similar result to that of Theorem 4 the case when the number of jobs and servers are equal and we can also prove a similar result corresponding to that of Theorem 5.

Theorem 8. *Let us consider $x \sim \mathcal{P}$. Conditioned on x, we study the setup where any set of x servers is equally likely to communicate with the master. Then the pairwise balanced job (n,k,k,n) assignment schemes (if exists) attain the least variance on the number of distinct jobs received at master amongst all balanced (n,k,k,n) assignment schemes.*

Theorem 9. *Let us consider $x \sim \mathcal{P}$. Conditioned on x, we study the setup where any set of x servers is equally likely to communicate with the master. Then both the pairwise heavy imbalanced job (n,k,r,c) assignment scheme (if exists) attain the largest variance on the number of distinct jobs received at master amongst all balanced (n,k,r,c) assignment schemes.*

VII. Proof of Corollary 1

Corollary 1. *For $x > c-r$, the expression of $\sigma_{D,x}(d)$ in Equation (28) in Theorem 2 goes to zero.*

Proof. Recall the expression of $\sigma_{D,x}(d)$ from Equation (28). Observe that expression $g(m+1,x)-g(m,x)$ from Equation (23) would be $\binom{c-2}{x-1}$ for $x > c-r$ as the second and third term in equation (23) goes to zero since $p \leq r$ and $x > c-r$.

$$g(m+1,x) - g(m,x) = \binom{c-2}{x-1} \tag{36}$$

Let us now compute $g(0,x)$ using the expression in (18) and (17) for $x > c-r$.

$$g(0,x) = \sum_{i=2}^{r+1}\sum_{j=2}^{r+1}(i-1)(j-1)\binom{c-2r}{x-i-j} = r^2\binom{c-2}{x-2} - 2r\binom{c-1}{x-1} + \binom{c}{x} \tag{37}$$

Thus, from equations (36) and (37), we get

$$g(m,x) = r^2\binom{c-2}{x-2} - 2r\binom{c-1}{x-1} + \binom{c}{x} + m \times \binom{c-2}{x-1} \tag{38}$$

Since, $x > c-r$, we may claim that the term $T_2(n,k,r,c)$ in Equation (28) in Thoerem 2 goes as follows.

$$T_2(n,k,r,c) = \sum_{t=1}^{r-1} t^2 \binom{r}{(t+1)}\binom{(c-r)}{(x-t-1)} = \left(r(r-1)\binom{c-2}{x-2} - r\binom{c-1}{x-1} + \binom{c}{x}\right) \tag{39}$$

Also observe that since $x > c-r$ the term $\binom{c-r}{x}$ goes to zero, hence not written in equation (15). Thus the numerator of the first term in equation (28) in Thoerem 2 is given by (from equations (37) and (38) and (39))

$$\begin{aligned}
&2.\sum_{m=0}^{r} \mathfrak{m}^D(m) g(m,x) + n\sum_{t=1}^{r-1} t^2 \binom{r}{(t+1)}\binom{(c-r)}{(x-t-1)} \\
&= \sum_{m=0}^{r}\left(2\mathfrak{m}^D(m)\left(r^2\binom{c-2}{x-2} - 2r\binom{c-1}{x-1} + \binom{c}{x}\right) + 2m\mathfrak{m}^D(m)\binom{c-2}{x-1}\right) \\
&\qquad\qquad + n\left(r(r-1)\binom{c-2}{x-2} - r\binom{c-1}{x-1} + \binom{c}{x}\right) \\
&\overset{(a)}{=} \left(r^2\binom{c-2}{x-2} - 2r\binom{c-1}{x-1} + \binom{c}{x}\right)n(n-1) + \binom{c-2}{x-1}ck(k-1) \\
&\qquad\qquad + n\left(r(r-1)\binom{c-2}{x-2} - r\binom{c-1}{x-1} + \binom{c}{x}\right) \\
&\overset{(b)}{=} \binom{c-2}{x-2}(nr(nr-1)) + \binom{c-2}{x-1}ck(k-1) - \binom{c-1}{x-1}(nr(2n-1)) + n^2\binom{c}{x} \\
&\overset{(c)}{=} \binom{c-2}{x-2}n^2r^2 + \binom{c-2}{x-1}ck^2 - \binom{c-1}{x-1}(nr(2n)) + n^2\binom{c}{x} \\
&\overset{(d)}{=} \binom{c-1}{x-1}nr \times kx - \binom{c-1}{x-1}(nr(2n)) + n^2\binom{c}{x} \\
&\overset{(e)}{=} \binom{c}{x}\left(\left(\frac{nrx}{c}\right)^2 - 2n\left(\frac{nrx}{c}\right) + n^2\right) \\
&\overset{(f)}{=} \binom{c}{x}\left(n \times \left(\frac{rx}{c} - 1\right)\right)^2
\end{aligned} \tag{40}$$

We now argue for each of the steps below.

- (a) follows since $\sum_{m=0}^{r} m \times \mathfrak{m}^D(m) = c\binom{k}{2}$ and $\sum_{m=0}^{r} \mathfrak{m}^D(m) = \binom{n}{2}$ in Equations (25) and (26)
- (b) follows by combining the coefficints of $\binom{c-2}{x-2}$, $\binom{c-1}{x-1}$ and $\binom{c}{x}$.
- (c) follows as $nr\binom{c-2}{x-2} + kc\binom{c-2}{x-1} = nr\binom{c-1}{x-1}$. This can be explained by the fact that $n \times r = k \times c$.
- (d) follows from the following set of equalities

$$\begin{aligned}
\binom{c-2}{x-2}n^2r^2 + \binom{c-2}{x-1}ck^2 = \frac{(c-1)!}{(x-2)!}(c-x-1)!\left(\frac{ck}{c-x} + \frac{k}{x-1}\right) &= \frac{(c-1)! \times nr \times kx(c-1)}{(x-2)!(c-x)(x-1)} \\
&= nr \times kx\binom{c-1}{x-1}
\end{aligned}$$

- (e) and (f) follow from the fact that $n \times r = k \times c$

Now, observe the second term of $\sigma_{D,x}(d)$ in equation (28) and we see that $T_1(n,k,r,c) = n \times \left(\frac{rx}{c} - 1\right)$ as $x > c-r$. Thus, using equation (40), we can say that $\sigma_{D,x}(d) = 0$ for $x > c-r$.

■

VIII. Proof of Theorem 3

To prove theorem 3, we first show some results using the convexity property of $g(.,x)$ [for a given x] as defined in Theorem 2. Note that $g(m+1,x)-g(m,x)$ as defined in Theorem 2 increases with p.

Claim 1. *Recall the definition of $g(p,x)$ as defined in Equation (16). Then, the following can be said (for every $c-2r+1 \geq x > 1$):*

- $\frac{g(m+k_1,x)-g(m,x))}{k_1} > \frac{g(m,x))-g(m-k_2,x)}{k_2}$ $\forall$ $k_1, k_2 \in \mathbb{N}$.
- $\frac{g(m+k_1,x)-g(m,x)}{k_1} > \frac{g(m+1,x))-g(m-k_2+1,x)}{k_2}$ $\forall$ $k_1, k_2 \in \mathbb{N}$ *with* $(k_1,k_2) \neq (1,1)$.
- $\frac{g(m+k_1+1,x)-g(m+1,x)}{k_1} > \frac{g(m,x)-g(m-k_2,x)}{k_2}$ $\forall$ $k_1, k_2 \in \mathbb{N}$ *with* $(k_1,k_2) \neq (1,1)$.
- $\frac{g(m+k_1,x)-g(m,x)}{k_1} > \frac{g(m+k_2,x)-g(m,x)}{k_2}$ $\forall$ $k_1, k_2 \in \mathbb{N}, k_1 > k_2$.

However, the inequalities may not strict for $x > c-2r+1$ but maybe attained with equality

Proof. We first prove the strict inequalities assuming $x \leq c-2r+1$

$$\frac{g(m+k_1,x)-g(m,x)}{k_1} = \frac{\sum_{i=0}^{i=k_1-1}(g(m+i+1,x)-g(m+i,x))}{k_1} \overset{(a)}{\geq} \frac{k_1(g(m+1,x)-g(m,x)}{k_1} \geq g(m+1,x)-g(m,x) \tag{41}$$

Now consider,

$$\frac{g(m+k_1+1,x)-g(m+1,x)}{k_1} = \frac{\sum_{i=1}^{i=k_1}(g(m+i+1,x)-g(m+i,x))}{k_1} \overset{(d)}{\geq} \frac{k_1(g(m+2,x)-g(m+1,x)}{k_1} > g(m+2,x)-g(m+1,x) \tag{42}$$

Note (a) and (d) follow since since $g(m+i+1,x)-g(m+i,x) > g(m+1,x)-g(m,x)$ $\forall i \in \mathbb{N}$ as shown in in Claim 2 where

$$g(m,x)-g(m-1,x) = \left[\binom{(c-2)}{(x-1)} - 2\binom{(c-r-1)}{(x-1)} + \binom{(c-2r+m-1)}{(x-1)}\right]$$

strictly increases with m for every $c-2r+1 > x > 1$. Also observe that (a) and (d) would be tight inequalities when $k_1 > 1$.

Similarly,

$$\frac{g(m,x)-g(m-k_2,x)}{k_2} = \frac{\sum_{i=0}^{i=k_2-1}(g(m-i,x)-g(m-i-1,x))}{k_2} \overset{(b)}{\leq} \frac{k_2(g(m,x)-g(m-1,x)}{k_2} \leq g(m,x)-g(m-1,x) \tag{43}$$

Similarly,

$$\frac{g(m+1,x)-g(m-k_2+1,x)}{k_2} = \frac{\sum_{i=0}^{i=k_2-1}(g(m-i+1,x)-g(m-i,x))}{k_2} \overset{(c)}{\leq} \frac{k_2(g(m+1,x)-g(m,x)}{k_2} \leq g(m+1,x)-g(m,x) \tag{44}$$

Note (b) and (c) follow since $g(m-i,x)-g(m-i-1,x) \leq g(m,x)-g(m-1,x)$ $\forall i \in \mathbb{N}$ as shown in Claim 2 where $g(m+1,x)-g(m,x)$ strictly increases with m for every $c-2r+1 > x > 1$. Also observe that (b) and (c) would be tight inequalities when $k_2 > 1$.

Thus we can say from Equations (41),(42),(43),(44) and the fact that $g(m+1)-g(m)$ strictly increases with m, the following equations (45),(46),(47).

$$\frac{g(m+k_1,x)-g(m,x)}{k_1} > \frac{g(m,x)-g(m-k_2,x)}{k_2} \forall k_1,k_2 \in \mathbb{N} \tag{45}$$

and

$$\frac{g(m+k_1,x)-g(m,x)}{k_1} > \frac{g(m+1,x)-g(m-k_2+1,x)}{k_2} \forall k_1,k_2 \in \mathbb{N} \text{ with } (k_1,k_2) \neq (1,1). \tag{46}$$

$$\frac{g(m+k_1+1,x)-g(m+1,x)}{k_1} > \frac{g(m+1,x)-g(m-k_2+1,x)}{k_2} \forall k_1,k_2 \in \mathbb{N} \text{ with } (k_1,k_2) \neq (1,1). \tag{47}$$

Thus the first, second and third inequalities in Claim 1are precisely the equations (45),(46) and (47). Now consider the following for $k_1 > k_2$

$$\frac{g(m+k_1,x)-g(m,x)}{k_1} \geq \frac{g(m+k_2,x)-g(m,x)}{k_2} \Leftrightarrow \frac{g(m+k_1,x)-g(m+k_2,x)}{k_1-k_2} \geq \frac{g(m+k_2,x)-g(m,x)}{k_2}$$

Note that since $k_1 > k_2$, we prove the inequality on R.H.S from Equation (45). Thus, the fourth inequality in Claim 1 holds true.

Observe that for $x > c-2r+1$, the function $g(m+1,x)-g(m,x)$ is not strictly increasing with m as the last term $\binom{c-2r+m-1}{x-1}$ maybe zero when m equals 0 or 1 and thus the inequalities in the lemma maybe attained with equality. ■

Let us now re-state and prove Theorem 3.

Theorem. *The least variance of the number of distinct jobs received at the master for any $x \in [2, c-2r+1]$ is attained uniquely by the pairwise balanced job (n,k,r,c) assignment schemes (if it exists) amongst all balanced (n,k,r,c) assignments.*

However, for $x > c-2r+1$, the result on least variance still holds but the uniqueness is not guaranteed.

Recall the expression of the variance from Theorem 2 and observe that $g(m,x)$ is a convex function in m. The essential idea of the proof is to use the convexity property of $g(.,x)$ and show that this expression takes the least value when $\mathfrak{m}^D(m)$ is non-zero for exactly two consecutive values of m.

Proof. Recall the definition of $\mathfrak{m}^D(m)$ from Equation 24 which denotes the number of pairs of jobs that are assigned together to exactly m servers.

Now we know that $\mathfrak{m}^D(m) = 0$ only for $m \neq l, l+1$ for any pairwise balanced job (n,k,r,c) assignment scheme. This would ensure that $l = \lfloor \frac{c.\binom{k}{2}}{\binom{n}{2}} \rfloor$ as shown in Lemma 3.

Let us consider another job assignment D_1 which is a balanced (n,k,r,c) assignment scheme.

First observe that $\sum_{m=0}^{r} \mathfrak{m}^D(m) = \sum_{m=0}^{r} \mathfrak{m}^{D_1}(m) = \binom{n}{2}$ which follows from Equation (25) and $\sum_{m=0}^{r} m \times \mathfrak{m}^D(m) = \sum_{m=0}^{r} m \times \mathfrak{m}^{D_1}(m) = c\binom{k}{2}$ follows from Equation (26).

We thus have

$$\sum_{m=0}^{r} \mathfrak{m}^D(m) = \sum_{m=0}^{r} \mathfrak{m}^{D_1}(m) = \binom{n}{2} \text{ and } \sum_{m=0}^{r} m \times \mathfrak{m}^D(m) = \sum_{m=0}^{r} m \times \mathfrak{m}^{D_1}(m) = c\binom{k}{2} \tag{48}$$

We now consider four different cases and in each of these cases, we show that the variance is the least for the assignment D. Also, we first consider the case where $x \leq c-2r+1$ to prove that least variance is uniquely attained by pairwise balanced designs.

- Case 1: $\mathfrak{m}^{D_1}(l) \geq \mathfrak{m}^{D}(l)$ but $\mathfrak{m}^{D_1}(l+1) \geq \mathfrak{m}^{D_1}(l+1)$
Note that $\mathfrak{m}^{D_1}(m) \geq \mathfrak{m}^{D}(m) \forall m \in [r]$ which follows since $\mathfrak{m}^{D}(m) = 0 \forall m \neq l, l+1$. However, equation (48) would imply that $\mathfrak{m}^{D}(m) = \mathfrak{m}^{D_1}(m) \ \forall \ m \in [r]$ which would imply distribution D_1 has same variance of distinct jobs as that of distribution D which follows from Equation 28 in Theorem 2 as the variance $\sigma_{D,x}$ is just a function of $\mathfrak{m}^{D}(.)$ other than design parameters n, k, r and c.
- Case 2: $\mathfrak{m}^{D_1}(l) < \mathfrak{m}^{D}(l)$ but $\mathfrak{m}^{D_1}(l+1) \geq \mathfrak{m}^{D_1}(l+1)$
Let us denote $x_m = \mathfrak{m}^{D_1}(m) - \mathfrak{m}^{D}(m) \ \forall m \in [0, r]$. Clearly, $x_m < 0$ only for $m = l$ as for every $m \neq l, l+1$ we have $\mathfrak{m}^{D}(m) = 0$
Observe that equation (48) ensures that $\sum_{m=0}^{r} x_m = \sum_{m=0}^{r} m \times x_m = 0$.
Let us denote $\sum_{m=0}^{l-1} x_m = x$ and $\sum_{m=l+1}^{r} x_m = y$. Since $\sum_{m=0}^{r} x_m = 0$, we can say that $x_l = -(x+y)$.
Thus,

$$\sum_{m=0}^{r} m.x_m = 0 \Leftrightarrow (x+y)l = \sum_{m=l+1}^{r} m \times x_m + \sum_{m<l} m \times x_m \Leftrightarrow \sum_{m=l+1}^{r} x_m \times (m-l) = \sum_{q=0}^{l-1} x_q \times (l-q) \quad (49)$$

However, we know from Claim 1 that

$$\frac{g(m,x) - g(l,x)}{m-l} > \frac{g(l,x) - g(q,x)}{l-q} \forall m > l > q \quad (50)$$

These equations (49) and (50) and the fact that $x_m \geq 0$ for $m \neq l, m \in [0, r]$ above would imply:

$$\begin{aligned} \sum_{m=l+1}^{r} x_m(g(m,x) - g(l,x)) > \sum_{q=0}^{l-1} x_q(g(l,x) - g(q,x)) &\overset{(c)}{\Leftrightarrow} \sum_{\substack{m \neq l \\ m \in [0,r]}} x_m.g(m,x) - \sum_{\substack{m \neq l \\ m \in [0,r]}} x_m g(l,x) > 0 \\ &\Leftrightarrow \sum_{\substack{m \neq l \\ m \in [0,r]}} x_m.g(m,x) + x_l.g(l,x) > 0 \\ &\Leftrightarrow \sum_{m=0}^{r} x_m g(m,x) > 0 \\ &\overset{(f)}{\Leftrightarrow} \sum_{m=0}^{r} \mathfrak{m}^{D_1}(m) g(m,x) > \sum_{m=0}^{r} \mathfrak{m}^{D}(m) g(m,x) \end{aligned} \quad (51)$$

Note (c) follows since $\sum_{m=0}^{r} x_m = 0$ and (f) follows from the fact that $x_m = \mathfrak{m}^{D_1}(m) - \mathfrak{m}^{D}(m)$
Now let us consider the numerator of the first term in $\sigma_{D,x}(d)$ as in theorem 28 which can be written as $2. \sum_{m=0}^{r} \mathfrak{m}^{D}(m) g(m,x) + n. \left(\sum_{t=1}^{r-1} t^2 \binom{r}{(t+1)} \binom{(c-r)}{(x-t-1)} \right)$.
Thus the inequality proven in equation (51) would imply that distribution D_1 has higher variance of number of distinct jobs received than that of distribution D.
- Case 3: $\mathfrak{m}^{D_1}(l+1) < \mathfrak{m}^{D}(l+1)$ but $\mathfrak{m}^{D_1}(l) \geq \mathfrak{m}^{D}(l)$
Note this can be proven in a very similar way as that of Case 2. The entire proof could be done for $l+1$ instead of l
- Case 4: $\mathfrak{m}^{D_1}(l+1) < \mathfrak{m}^{D}(l+1)$ and $\mathfrak{m}^{D_1}(l) < \mathfrak{m}^{D}(l)$
Let us denote $x_m = \mathfrak{m}^{D_1}(m) - \mathfrak{m}^{D}(m) \ \forall m \in [0, r]$. Clearly, $x_m < 0$ only for $m = l, l+1$.
Now equation (48) ensures that $\sum_{m=0}^{r} m \times x_m = 0$ and $\sum_{m=0}^{r} x_m = 0$.

Let us denote $\sum_{m=0}^{l-1} x_m = \alpha_1 + \alpha_2$ and $\sum_{m=l+2}^{r} x_m = \beta_1 + \beta_2$. Since $\sum_{m=0}^{r} x_m = 0$, we can say that

$$x_l = -(\alpha_1 + \beta_1) \text{ and } x_{l+1} = -(\alpha_2 + \beta_2) \text{ for some } \alpha_1, \beta_1, \alpha_2, \beta_2 \in \mathbb{N} \tag{52}$$

Now define $[y_m]_{m=0}^{r}$ and $[z_m]_{m=0}^{r}$as follows below.

- $y_m = x_m \times \frac{\alpha_1}{\alpha_1+\alpha_2}$ and $z_m = x_m \times \frac{\alpha_2}{\alpha_1+\alpha_2}$ if $0 \leq m < l$
- $y_m = x_m \times \frac{\beta_1}{\beta_1+\beta_2}$ and $z_m = x_m \times \frac{\beta_2}{\beta_1+\beta_2}$ for $l+1 < m \leq r$
- $y_m = \alpha_1$ and $z_m = \alpha_2$ if $m = l$
- $y_m = \beta_1$ and $z_m = \beta_2$ if $m = l+1$

Thus,
$\sum_{m=l+2}^{r} y_m = \beta_1$; $\sum_{m=l+2}^{r} z_m = \beta_2$; $\sum_{m=0}^{l-1} y_m = \alpha_1$; $\sum_{m=0}^{l-1} z_m = \alpha_2$;
Thus $\sum_{m=0}^{r} m x_m = 0$ implies the following set of conditions.

$$(\alpha_1 + \beta_1)l + (\alpha_2 + \beta_2)(l+1) = \sum_{m=l+2}^{r} m \times x_m + \sum_{m=0}^{l-1} m \times x_m$$

$$\overset{(d)}{\Leftrightarrow} \left(\sum_{m=l+2}^{r} y_m + \sum_{m=0}^{l-1} y_m \right) l + \left(\sum_{m=l+2}^{r} z_m + \sum_{m=0}^{l-1} z_m \right) (l+1) = \sum_{m=l+2}^{r} m \times (y_m + z_m) + \sum_{m=0}^{l-1} m \times (y_m + z_m)$$

$$\Leftrightarrow \sum_{m=l+2}^{r} y_m(m-l) + \sum_{m=l+2}^{r} z_m(m-l-1) = \sum_{q=0}^{l-1} y_q(l-q) + \sum_{q=0}^{l-1} z_q(l+1-q) \tag{53}$$

Note (d) follows since

i) $y_m + z_m = x_m \forall m \in [0, l-1] \cup [l+2, r]$, ii) $\sum_{m=l+2}^{r} y_m + \sum_{q=0}^{l-1} y_q = (\alpha_1 + \beta_1)$ and

iii) $\sum_{m=l+2}^{r} z_m + \sum_{q=0}^{l-1} z_q = (\alpha_2 + \beta_2)$

Now we can say from Claim 1 that

$$\frac{g(m,x) - g(t,x)}{m-t} > \frac{g(u,x) - g(q,x)}{u-q} \;\forall m > l+1, q < l \text{ and } t, u \in \{l, l+1\}. \tag{54}$$

Thus we can say the following from equations (53) and (54) and the fact that $x_m, y_m, z_m > 0$ for $m \in [0, r]; m \neq \{l, l+1\}$.

$$\sum_{m=l+2}^{r} y_m(g(m,x)-g(l,x)) + \sum_{m=l+2}^{r} z_m(g(m,x)-g(l+1,x)) \tag{55}$$

$$> \sum_{q=0}^{l-1} y_q(g(l,x)-g(q,x)) + \sum_{q=0}^{l-1} z_q(g(l+1,x)-g(q,x)) \tag{56}$$

$$\overset{(e)}{\Leftrightarrow} \sum_{\substack{m\neq l,l+1\\ m\in[0,r]}} (y_m+z_m)g(m,x) + (\alpha_1+\beta_1)g(l,x) + (\alpha_2+\beta_2)g(l+1,x) > 0 \tag{57}$$

$$\overset{(g)}{\Leftrightarrow} \sum_{\substack{m\neq l,l+1\\ m\in[0,r]}} x_m.g(m,x) + x_l.g(l,x) + x_{l+1}.g(l+1,x) > 0 \tag{58}$$

$$\Leftrightarrow \sum_{m=0}^{r} x_m.g(m,x) > 0 \tag{59}$$

$$\overset{(h)}{\Leftrightarrow} \sum_{m=0}^{r} \mathfrak{m}^{D_1}(m)g(m,x) > \sum_{m=0}^{r} \mathfrak{m}^{D}(m)g(m,x) \tag{60}$$

Note (e) follows since $\sum\limits_{m=l+2}^{r} y_m + \sum\limits_{q=0}^{l-1} y_q = (\alpha_1+\beta_1)$ and $\sum\limits_{m=l+2}^{r} z_m + \sum\limits_{q=0}^{l-1} z_q = (\alpha_2+\beta_2)$ from Equation (52) and the fact $y_m + z_m = x_m \forall m \in [0, l-1] \cup [l+2, r]$.
(g) follows from the fact that $x_l = -(\alpha_1+\beta_1)$ and $x_{l+1} = -(\alpha_2+\beta_2)$ (h) follows from the fact that that $x_m = \mathfrak{m}^{D_1}(m) - \mathfrak{m}^{D}(m)$
Now let us consider the numerator of the first term in $\sigma_{D,x}(d)$ as in theorem 28 which can be written as $2.\sum\limits_{m=0}^{r} \mathfrak{m}^{D}(m)g(m,x) + n.\left(\sum\limits_{t=1}^{r-1} t^2 \binom{r}{(t+1)}\binom{(c-r)}{(x-t-1)}\right)$.
Thus the inequality proven in the previous equation (55) would imply that distribution D_1 has higher variance of number of distinct jobs received than that of distribution D.
Note that other than Case 1 in this proof (where we show D_1 is also a pairwise balanced assignment), the distribution D_1 is not a pairwise balanced job design and we show strict optimality of pairwise balanced job design over any other balanced assignment for $x \leq c-2r+1$. However, if $x \geq c-2r+1$, the inequalities still hold but are not strict as it follows from Claim 1.
Combining these statements, we prove the strict optimality (attains least variance)of pairwise balanced designs for $x \leq c-2r+1$ and weak optimality for $x > c-2r+1$

∎

IX. Proof of Theorem 5

Let us first restate and prove Theorem 5.

Theorem. *The largest variance of the number of distinct jobs d received at the master for any $x \in [2, c-2r+1]$ is attained uniquely by the pairwise balanced job (n,k,r,c) assignment schemes (if it exists) amongst all balanced (n,k,r,c) assignments.*

However, for $x > c-2r+1$, the result on the largest variance still holds but the uniqueness is not guaranteed.

Recall the expression of the variance from Theorem 2 and observe that $g(m,x)$ is a convex function in m. The essential idea of the proof is to use the convexity property of $g(.,x)$ and show that this expression takes the largest value when $\mathfrak{m}^{D}(m)$ are non-zero for exactly those values of m when it is zero or r.

Proof. Recall the definition of $\mathfrak{m}^D(m)$ from Equation 24 which denotes the number of pairs of jobs that are assigned together to exactly m servers.

Now we know that $\mathfrak{m}^D(m) = 0$ only for $m \neq 0, r$ for any pairwise balanced job (n,k,r,c) assignment scheme.

Let us consider another job assignment D_1 which is a balanced (n,k,r,c) assignment scheme.

First observe that $\sum\limits_{m=0}^{r} \mathfrak{m}^D(m) = \sum\limits_{m=0}^{r} \mathfrak{m}^{D_1}(m) = \binom{n}{2}$ which follows from Equation (25) and $\sum\limits_{m=0}^{r} m \times \mathfrak{m}^D(m) = \sum\limits_{m=0}^{r} m \times \mathfrak{m}^{D_1}(m) = c\binom{k}{2}$ follows from Equation (26).

We thus have

$$\sum_{m=0}^{r} \mathfrak{m}^D(m) = \sum_{m=0}^{r} \mathfrak{m}^{D_1}(m) = \binom{n}{2} \text{ and } \sum_{m=0}^{r} m \times \mathfrak{m}^D(m) = \sum_{m=0}^{r} m \times \mathfrak{m}^{D_1}(m) = c\binom{k}{2} \tag{61}$$

We now consider four different cases and in each of these cases, we show that the variance is the least for the assignment D. Also, we first consider the case where $x \leq c - 2r + 1$ to prove that the largest variance is uniquely attained by pairwise heavy imbalanced designs.

- Case 1: $\mathfrak{m}^{D_1}(r) \geq \mathfrak{m}^D(r)$
  Note that this implies that $\sum_{m=0}^{r} m \times \mathfrak{m}^{D_1}(m) \geq \sum_{m=0}^{r} m \times \mathfrak{m}^D(m)$ as $\mathfrak{m}^D(m) = 0 \forall m \in [0, r-1]$. Thus, equation (61) implies that $\mathfrak{m}^{D_1}(m) = \mathfrak{m}^D(m) \forall m \in [r]$ which would imply distribution D_1 has same variance of distinct jobs as that of distribution D which follows from Equation (28) in Theorem 2 as the variance $\sigma_{D,x}$ is just a function of $\mathfrak{m}^D(.)$ other than design parameters n, k, r and c.
- Case 2: $\mathfrak{m}^{D_1}(r) < \mathfrak{m}^D(r)$ but $\mathfrak{m}^{D_1}(0) \geq \mathfrak{m}^D(0)$
  Let us denote $x_m = \mathfrak{m}^{D_1}(m) - \mathfrak{m}^D(m)\ \forall m \in [0, r]$. Clearly, $x_m < 0$ only for $m = r$ as for every $m \neq r$ we have $\mathfrak{m}^D(m) = 0$.

  Observe that equation (48) ensures that $\sum\limits_{m=0}^{r} x_m = \sum\limits_{m=0}^{r} m \times x_m = 0$.

  Let us denote $\sum\limits_{m=1}^{r} x_m = x$. Since $\sum\limits_{m=0}^{r} x_m = 0$, we can say that $x_r = -x$.

  Thus,

$$\begin{aligned} \sum_{m=0}^{r} x_m \times g(m,x) &= \sum_{m=0}^{r-1} x_m \times g(m,x) - x \times g(r,x) \\ &= \sum_{m=0}^{r-1} x_m \times (g(m,x) - g(r,x)) \overset{(a)}{\leq} 0. \end{aligned} \tag{62}$$

  Note that (a) follows from the fact that $x_m > 0$ for all $m \in [0, r-1]$ and the fact that $g(m,x)$ is monotonic in m.
- Case 3: $\mathfrak{m}^{D_1}(0) < \mathfrak{m}^D(0)$ and $\mathfrak{m}^{D_1}(r) < \mathfrak{m}^D(r)$
  Let us denote $x_m = \mathfrak{m}^{D_1}(m) - \mathfrak{m}^D(m)\ \forall m \in [0, r]$. Clearly, $x_m < 0$ only for $m = 0, r$.

  Now equation (61) ensures that $\sum\limits_{m=0}^{r} m \times x_m = 0$ and $\sum\limits_{m=0}^{r} x_m = 0$ Observe that $x_0 + x_r = -\sum\limits_{m=1}^{r-1} \times x_m$

  Now define $[y_m]_{m=0}^{r}$ and $[z_m]_{m=0}^{r}$as follows below.
  - $y_m = x_m \times \frac{x_0}{x_0 + x_r}$ and $z_m = x_m \times \frac{x_r}{x_0 + x_r}$ if $0 < m < r$

  Also note that $y_m + z_m = x_m \forall m \in [1, r-1]$ and also $\sum_{l=1}^{r-1} y_l = -x_0$ and $\sum_{l=1}^{r-1} z_l = -x_r$. Thus, $\sum\limits_{m=0}^{r} m x_m = 0$ implies

$$\sum_{m=1}^{l-1} (r-m) z_m = \sum_{q=1}^{l-1} q \times y_q \tag{63}$$

Now we can say from Claim 1 that

$$\frac{g(r,x)-g(t,x)}{r-t} > \frac{g(t,x)-g(0,x)}{t-0} \ \forall t \in [0,r]. \tag{64}$$

Thus we can say the following from equations (63) and (64) and the fact that $x_m, y_m, z_m > 0$ for $m \in [0,r]$;

$$\begin{aligned}
&\sum_{m=1}^{r-1} z_m(g(r,x)-g(m,x)) > \sum_{q=1}^{r-1} y_q(g(q,x)-g(0,x)) \\
\Leftrightarrow &(\sum_{m=1}^{r-1} z_m)g(r,x) - \sum_{m=1}^{r-1}(y_m+z_m)g(m,x) + (\sum_{m=0}^{r-1} y_m)g(0,x) > 0 \\
\overset{(b)}{\Leftrightarrow} &- x_r g(r,x) - \sum_{m=1}^{r-1} x_m g(m,x) - x_0 g(0,x) > 0 \\
\Leftrightarrow &\sum_{m=0}^{r} x_m g(m,r) < 0 \\
\Leftrightarrow &\sum_{m=0}^{r} \mathfrak{m}^{D_1}(m)g(m,x) < \sum_{m=0}^{r} \mathfrak{m}^{D}(m)g(m,x)
\end{aligned} \tag{65}$$

Note (b) follows from $y_m + z_m = x_m \forall m \in [1, r-1]$ and also $\sum_{l=1}^{r-1} y_l = -x_0$ and $\sum_{l=1}^{r-1} z_l = -x_r$. Now let us consider the numerator of the first term in $\sigma_{D,x}(d)$ as in theorem 28 which can be written as $2.\sum_{m=0}^{r} \mathfrak{m}^D(m)g(m,x) + n.\left(\sum_{t=1}^{r-1} t^2\binom{r}{(t+1)}\binom{(c-r)}{(x-t-1)}\right)$.
Thus the inequality proven in the previous equation (55) would imply that distribution D_1 has a smaller variance of the number of distinct jobs received at the master than that of distribution D, thus proving our desired result.
Note that other than Case 1 in this proof (where we show D_1 is also a pairwise heavy imbalanced job assignment), the distribution D_1 is not a pairwise heavy imbalanced job assignment and we show strict optimality of pairwise balanced job design over any other balanced assignment for $x \leq c-2r+1$. However, if $x \geq c-2r+1$, the inequalities still hold but are not strict as it follows from Claim 1. Combining these statements, we prove the strict optimality (attains least variance)of pairwise balanced designs for $x \leq c-2r+1$ and weak optimality for $x > c-2r+1$

■

X. Proof of Theorem 4

To prove this theorem, we first prove Claim 2 and 3. These claims would be used in the proof of Theorem 4.

We now define a random variable Y^D as follows for balanced (n,k,r,c) assignment D as the number of servers in which a pair of jobs chosen uniformly at random occur together. Formally, we can say that

$$\mathbb{P}[Y^D = p] = \frac{\mathfrak{m}^D(p)}{\binom{n}{2}} \text{for any integer } p \in [0,r] \tag{66}$$

Observe that it is a valid distribution as $\sum_{p=0}^{r} \mathfrak{m}^D(p) = \binom{n}{2}$ in Equation (25).

Claim 2. *For any balanced (n,k,r,c) assignment D, the variance of Y^D is linearly proportional to the variance of distinct jobs d received at master when any 2 servers chosen uniformly at random return i.e. are able to communicate their results to the master. We can also state it as follows.*

$$\sigma_{D,2}(d) = \frac{\binom{n}{2}.\sigma(Y^D) + \frac{\left(c\binom{k}{2}\right)^2}{\binom{n}{2}} + n.\left(\binom{r}{2}\right) - c.\left(\binom{k}{2}\right)}{\binom{c}{2}} - \left(\frac{n\binom{r}{2}}{\binom{c}{2}}\right)^2 \tag{67}$$

Observe for $n = c$, we can say that $\sigma(Y^D) = \sigma_{D,2}(d)$

Proof. Consider the numerator of first term in $\sigma_{D,x}(d)$ in Equation (28) which had been shown to be $2.\sum_{m=0}^{r} \mathfrak{m}^D(m)g(m,x) + n.\left(\sum_{t=1}^{r-1} t^2\binom{r}{(t+1)}\binom{(c-r)}{(x-t-1)}\right)$.

Note that here we consider $x = 2$ as we consider the case when any 2 servers return.

Thus the first term in numerator of $\sigma_{D,2}(d)$ becomes $2.\sum_{m=0}^{r} g(m,x) + n.\left(\sum_{t=1}^{r-1} t^2\binom{r}{(t+1)}\binom{(c-r)}{(1-t)}\right)$.

Now, observe that $g(0,x) = \sum_{i=2}^{r+1}\sum_{j=2}^{r+1}(i-1).(j-1).\binom{(c-2r)}{(x-i-j)}.\binom{r}{i}.\binom{r}{j}$ from Equation (17). For $x = 2$, we can show that it goes to 0 as $g(m,2) - g(m-1,2)$ reduces to $[(n-2) - 2(n-k-1) + (n-2k+m-1)] = m-1$ from Equation (23) in Theorem 2. Thus we can argue that $g(m,2) = 1+2+\ldots+m-1 = \frac{m(m-1)}{2}$.

Thus for $x = 2$,

$$\begin{aligned}
&2.\sum_{m=0}^{r} \mathfrak{m}^D(m)g(m,2) + n.\left(\sum_{t=1}^{k-1} t^2\binom{r}{(t+1)}\binom{(c-r)}{(2-t-1)}\right) \\
&= \sum_{m=0}^{r} \mathfrak{m}^D(m)m(m-1) + n.\binom{r}{2} \\
&\overset{(a)}{=} \sum_{m=0}^{r} m^2\mathfrak{m}^D(m) + n\binom{r}{2} - c\binom{k}{2}
\end{aligned}$$

(a) follows since $\sum_{m=0}^{r} m\mathfrak{m}^D(m) = c\binom{k}{2}$ in Equation (26).

Now consider

$$\begin{aligned}
\sigma_{D,2}(d) &= \frac{2.\sum_{m=0}^{r} \mathfrak{m}^D(m)g(m,2) + n.\left(\sum_{t=1}^{k-1} t^2\binom{k}{(t+1)}\binom{(n-k)}{(1-t)}\right)}{\binom{c}{2}} - \left(\frac{n\sum_{t=1}^{r-1} t\binom{r}{(t+1)}\binom{(c-r)}{(1-t)}}{\binom{c}{2}}\right)^2 \\
&= \frac{\sum_{m=0}^{r} m^2.\mathfrak{m}^D(m) + n.\binom{r}{2} - c.\binom{k}{2}}{\binom{c}{2}} - \left(\frac{n\binom{r}{2}}{\binom{c}{2}}\right)^2
\end{aligned} \tag{68}$$

We can observe that $\mathbb{E}[Y^D] = \frac{\sum_{m=0}^{r} m\mathfrak{m}^D(m)}{\binom{n}{2}} = \frac{c\binom{k}{2}}{\binom{n}{2}}$ as $\sum_{m=0}^{r} m\mathfrak{m}^D(m) = c\binom{k}{2}$ [equation (26)] using the definition of Y^D in Equation (66).

$$\begin{aligned}
\sigma(Y^D) &= \mathbb{E}[(Y^D)^2] - (\mathbb{E}[Y^D])^2 \\
&= \frac{1}{\binom{n}{2}}\sum_{m=0}^{r} m^2\mathfrak{m}^D(m) - \left(\frac{c\binom{k}{2}}{\binom{n}{2}}\right)^2
\end{aligned} \tag{69}$$

Thus, using Equations (69) and (68), we prove Claim 2.

∎

We now prove a lemma showing that the set of pairwise balanced job assignments is identical to the set of pairwise balanced server assignments when the number of jobs and the servers are identical.

Claim 3. *Amongst balanced (n, k, k, n) assignments, every pairwise balanced (n, k, k, n) job assignments is also a pairwise balanced server assignment and vice-versa.*

Proof. Consider the set of balanced (n, k, k, n) assignment schemes. Now recall the definition of y from Claim 2 where y denoted the number of servers where a pair of jobs chosen uniformly at random occurs together and for $n = c$ and $k = r$, we have

$$\sigma_{D,2}(d) = \sigma(Y^D). \tag{70}$$

Now we can compute that the $\mathbb{E}[Y^D] = \frac{n\binom{k}{2}}{\binom{n}{2}}$ which follows since every server has exactly $\binom{k}{2}$ pairs of jobs since each server is balanced (and is assigned k jobs) and we have exactly n servers. However, we have exactly $\binom{n}{2}$ total number of pairs of jobs hence this equality of expectation on y holds true. Also, observe from Theorem 6 (under $x = 2, n = c$ and $k = r$) that

$$\mathbb{E}_{D,2}[d] = n\left(1 - \frac{k(k-1)}{n(n-1)}\right) = \left(2k - \frac{k(k-1)}{n(n-1)}\right) = \mathbb{E}_D[2k - Y^D] \tag{71}$$

We first show that every pairwise and balanced job assignment is also a pairwise balanced server assignment scheme.

Suppose not and consider a balanced assignment scheme D which is pairwise job balanced but not pairwise server balanced. Now let us consider the scenario where $x = 2$ (exactly 2 randomly chosen servers) are able to communicate with the master. Now suppose D is not a pairwise balanced server design which implies that d (the number of distinct jobs received) can take at least 2 distinct integral values when $x = 2$ which are non-consecutive, hence $2k - d$ also has a support of at least 2 distinct non-consecutive integral values. However the random variable Y^D has a support of at most 2 over two consecutive indices(since it is a pairwise job-balanced design). Now observe that random $2k - d$ and Y^D have the same expectation (shown above in Equation (71)). Thus, the variance of $2k - d$ is clearly more than that of Y^D which is a contradiction from Equation (70).

Now we consider the other case where a balanced assignment scheme is a pairwise server balanced but not pairwise job balanced. Consider the scenario where $x = 2$ (exactly 2 randomly chosen servers) are able to communicate with the master. Since the number of distinct jobs received at the master can take exactly 2 values (as it is pairwise server balanced), thus, the random variable $2k - d$ has a support of 2 over two consecutive indices. Suppose the assignment scheme is not pairwise job balanced hence Y^D has a support of at least 2 elements which are non consecutive. Now observe that random variables Y^D and $2k - d$ have the same expectation in (71), hence the variance of Y^D has to be clearly larger than that $2k - d$ which is a contradiction from Equation (70).

Thus, we show that the set of balanced balanced job designs is identical to the set of pairwise balanced server designs.

■

Now we restate and prove Theorem 4.

Theorem. *The least variance of the number of distinct jobs received at the master for any $x \in [1, c-2r+1]$ is attained uniquely by both the pairwise balanced server (n, k, k, n) assignment schemes and pairwise balanced job (n, k, k, n) assignment schemes (if it exists) amongst all balanced (n, k, k, n) assignments.*

Proof. This directly follows from Theorem 3 and Claim 3. ■

XI. Constructions of pairwise balanced designs:

In general, these types of constructions can be done using balanced incomplete block designs as explained in [1]. In these type of construction, treatments are divided into blocks so that each block has the same number of treatment and each treatment occurs exactly the same number of times. However in balanced block constructions we also ensure that every pair of treatments occur together in the same number of blocks. Using theorem 3 we can say that it would have the least variance amongst distributions satisfying Definition 2 for a given n, k, r and c.

Various methodologies of balanced block constructions have been proposed in [1], [3]. The famous Bruck-Ryser-Chowla theorem gives some necessary condition on n, k, c, r so that it might be possible to have a balanced symmetric design.

We discuss one construction from [1] using vector spaces.

A. Construction using vector sub-spaces

Let us discuss such a construction. Choose a finite field vector space of dimension N, Let us denote each job as the subspaces of this vector space of dimensionality $K = 1$. Also we denote each server as a subspace of dimensionality C ($C > 1$). Now only those jobs are present in a server such that the vector-space corresponding to a job is the subspace of the vector-space denoting the corresponding server.

Note that the number of jobs and server is same since number of subspaces of dimensionality $N - K$ is same as the number of subspaces of dimensionality K. Also by geometry we can show that each sub-space of dimensionality K has a fixed number of sub-spaces of dimensionality $N - K$. Thus both conditions in Definition 2 are satisfied.

In this construction, every pair of jobs must occur together in exactly the same number of servers and also every pair of servers have the same number of common jobs thus, it is both a pairwise balanced job and pairwise balanced and pairwise balanced server assignment scheme. Thus, according to Theorem 3, it attains the least variance on the distinct number of jobs received at the master.

However such constructions are only possible from $n = \frac{q^a-1}{q-1}$, $k = \frac{q^b-1}{q-1}$, $c = \begin{bmatrix} a \\ b \end{bmatrix}_q$ where $\begin{bmatrix} a \\ b \end{bmatrix}_q$ denotes the number of b-dimensional subspace of a-dimensional space on a finite field of order q where q is the power of a prime. This holds true for some positive integer a and q is a power of some prime number since the cardinality of finite fields can also be a power of some prime number.

B. Construction using 2-D spread of points

Note this construction is somewhat similar to the planar construction you had described for $n = 9, k = 3$ case. We can generalise it for any $n = a.b$ (using a field for $\mathbb{F}_a$), $k = a$ and $a < b$ such that a is a power of some prime number say p and b is a multiple of $\frac{a}{p}$.

We could also do a similar construction for $n = a.b$, $k = b$, $a > b$ such that a is a power of some prime number say p and b is a multiple of $\frac{a}{p}$.

Recall that these constructions were done by treating the jobs as points and servers were treated as lines. Since these lines are constructed on a finite field, no two lines would have more than one job common.

Thus every pair of servers intersect in at-most one point, thus theorem 3 would imply that it has the least variance for every x.

XII. Construction of heavy pairwise imbalanced job assignment schemes

We present a construction for the case when r divides c. We present an assignment scheme based on a repetition coding scheme. We divide the servers in c/r groups of r servers each. In the first group of r servers, we assign each such server jobs numbered from a_1 to a_k. In the next set of r servers, we assign it jobs numbered from a_{k+1} to a_{2k} and we repeat this assignment scheme till all servers are exhausted.

Formally, a server $s_{i\times r+j}$ is assigned jobs from $a_{k\times i+1}$ to $a_{k\times(i+1)}$ for every $i \in [0, \frac{c}{r} - 1]$ and every $j \in [0, r]$. Observe that such a design is balanced as every job is assigned to r servers and every server has precisely k jobs. Also observe that in this construction, a pair of jobs either occur together in r servers or do not occur together at all, thus presenting a construction of a heavy pairwise imbalanced job assignment scheme.

REFERENCES

[1] R. C. BOSE, "On the construction of balanced incomplete block designs," *Annals of Eugenics*, vol. 9, no. 4, pp. 353–399, 1939. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-1809.1939.tb02219.x

[2] C. J. Colbourn and J. H. Dinitz, *Handbook of Combinatorial Designs, Second Edition (Discrete Mathematics and Its Applications)*. Chapman & Hall/CRC, 2006.

[3] D. A. Sprott, "Balanced incomplete block designs and tactical configurations," *Ann. Math. Statist.*, vol. 26, no. 4, pp. 752–758, 12 1955. [Online]. Available: https://doi.org/10.1214/aoms/1177728433