Virtual machine (VM) improves data center’s efficiency and flexibility. VM placement needs to consider multiple server resource demands. Including memory, CPU, power consumption, etc. Can be modeled as Multidimensional Bin Packing (MBP) problem. Assume fixed-value resource demands – deterministic demands.

**Stochastic Demand Challenge**
- Some resource demands are bursty and time varying.
- Such as network bandwidth demand.
- Model such demand as stochastic process – stochastic demand.
- Service level agreement (SLA).
- Challenge: No fixed value can be used in placement calculation.

**Goal**
- Effective VM placement to minimize the number of servers.
- Consider multiple resource demands.
- Consider both deterministic and stochastic resource demands.
- Need to satisfy SLA for stochastic demands.

**Problem Formulation**
- Model demands by using $N(\mu(v_i), \sigma^2(v_i))$ normal distribution.
- Calculate equivalent demand based on SLA to quantify the stochastic demands:
  \[
  \text{Total Equivalent Demand} = \sum_{i \in D} \mu(v_i) + \Phi^{-1}(\alpha) \sum_{i \in D} \sigma^2(v_i)
  \]
- $\Phi^{-1}(\alpha)$ is the quantile of standard normal distribution $N(1,0)$ when SLA is equal to $\alpha$.
- Calculate server resource utilization:
  \[
  \begin{align*}
  &\text{Deterministic:} \quad \forall p \in P, R_p(s) = \frac{\sum_{i \in D} D_p(v_i)}{C_p(s)} \\
  &\text{Stochastic:} \quad \forall q \in Q, R_q(s) = \frac{\sum_{i \in D} \mu_q(v_i) + \Phi^{-1}(\alpha) \sum_{i \in D} \sigma_q^2(v_i)}{C_q(s)}
  \end{align*}
  \]
- MSVP formulation:
  \[
  \begin{align*}
  &\text{minimize } |S| \\
  &\text{s.t. } \forall p \in P, \forall s \in S, R_p(s) \leq 1 \\
  &\forall q \in Q, \forall s \in S, R_q(s) \leq 1
  \end{align*}
  \]

**Algorithm**
- Max-Min Multidimensional Stochastic Bin Packing (M^3SBP)
  - Step 1:
    - Find candidate VMs that can be potentially put in current server.
    - Employ Total Equivalent Demand for stochastic resources.
    - Record the minimum resource utilization.
    - If no VM is found, open a new server.
  - Step 2:
    - Compare the minimum utilizations recorded in step 1.
    - Choose the VM that maximizes the minimum utilization.

**Performance Evaluation**
- Simulation configurations:
  - VMs with 4 different resource demands.
  - Deterministic: memory, CPU and power consumption.
  - Stochastic: bandwidth.
- Standard deviation to represent burst level.
- 2000 VMs and 99.99% SLA.
- Compare M^3SBP with benchmark algorithms.

**Results**
- Number of servers used by Max-Min:
  - M^3SBP: 386 servers; 0% violated servers.
  - Max-Min: 387 servers; 0.26% violated servers.
- Number of servers used by Max-Min:
  - M^3SBP: < 1.04% of Violated Servers.
  - Max-Min: > 1.04% of Violated Servers.

**Conclusion**
- M^3SBP – An effective VM placement algorithm.
- Calculate total equivalent demands for stochastic resources.
- Maximize the minimum utilization among all server resources.
- Satisfy SLA requirement with fewest servers.