Markov Model Based Experiment Comparison

Swati Sharma and Alefiya Hussain
sswati@cse.iitd.ernet.in, hussain@isi.edu

Objective

Ability to compare consecutive experiment runs – configuration & output data.

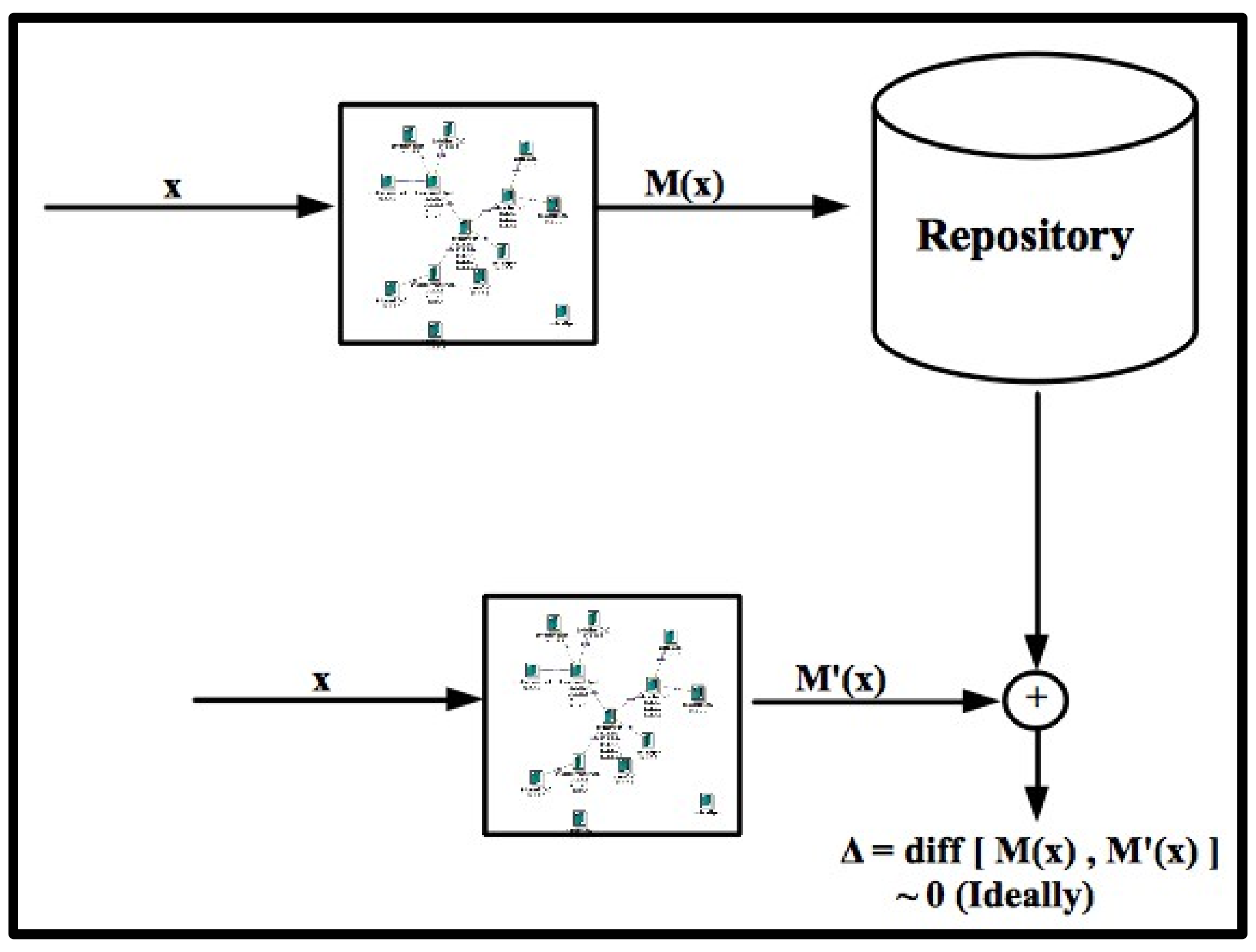


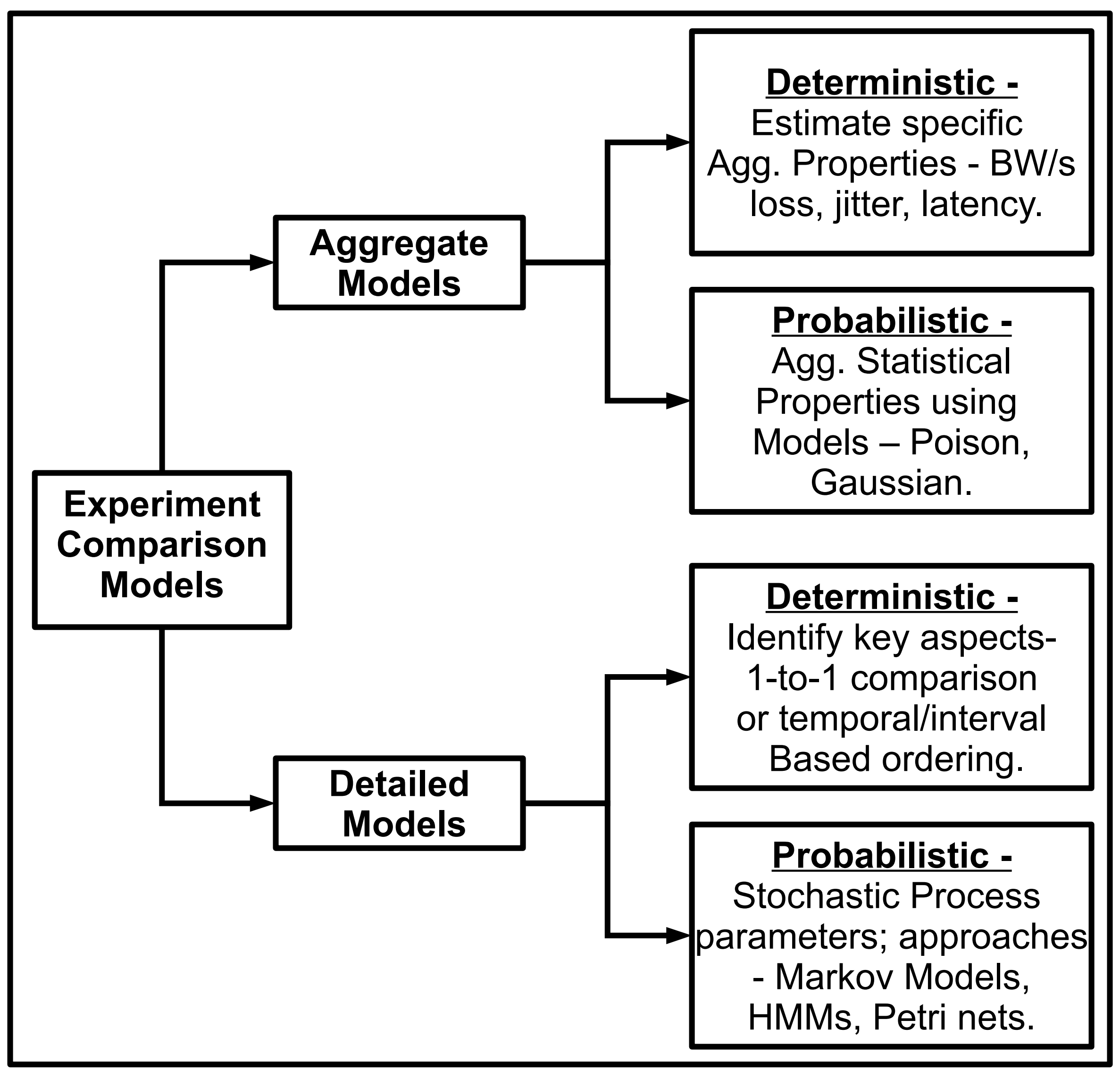
Figure 1 – Illustration of expt. comparison concept.

x: expt. config., M(x): comparison model

Motivation

- **Experiment components -**
  - **Deterministic** – simple computer programs.
  - **Non-Deterministic** – dynamic n/w behavior.
  - **Opportunistic** – attack models.
- **High-level aggregate metrics -**
  - Fail to capture complex configuration dependent dynamics.

Fundamentals



Experiment Comparison Models

Aggregate Models

Deterministic - Estimate specific Agg. Properties - BW/s loss, jitter, latency.

Probabilistic - Agg. Statistical Properties using Models – Poison, Gaussian.

Detailed Models

Deterministic - Identify key aspects- 1-to-1 comparison or temporal/interval Based ordering.

Probabilistic - Stochastic Process parameters; approaches - Markov Models, HMMs, Petri nets.

Figure 2 – Categorization of possible approaches.

1st-Order Markov Model

$$M(x)=(S, IS, A), \quad IS \subset S. \qquad (1)$$

$$P[q_t=S_j, q_{(t-1)}=S_i, q_{(t-2)}=S_k, ...] \; P[q_t=S_j, q_{(t-1)}=S_i]. \qquad (2)$$

- Eqn (1) : M(x) = Markov Model, 'S' = finite set of states, 'IS' = set of initial states, 'A' = Transition Prob. Matrix.
- Eqn (2) : M(x) = Sequence of stochastic events; state -
  - Dictated only by previous state.
  - Independent of path followed.

$$a_{i,j}=P[q_t=S_j, q_{(t-1)}=S_i], 1 \le i, j \le N, \quad a_{i,j} \ge 0, \sum_{i=1}^{N} a_{i,j}=1. \qquad (3)$$

- Eqn (3) : sum (all probabilities from a state) = 1.

Model Creation

- Obtain S (distinct minimal N-tuple packets), IS & A.
- Populate state transition diagram, save model.

Model Comparison

- Create model from several runs - ensure statistical soundness.
- Generate M'(x), find δ (degree of variability b/w experiment runs) – (4).

$$\delta=\sum_{i,j}(|a_{i,j}(M(x))-a_{i,j}(M'(x))|^2). \qquad (4)$$

- Lower δ → closer match between experiment runs.

Results - Experiment & Sample Data

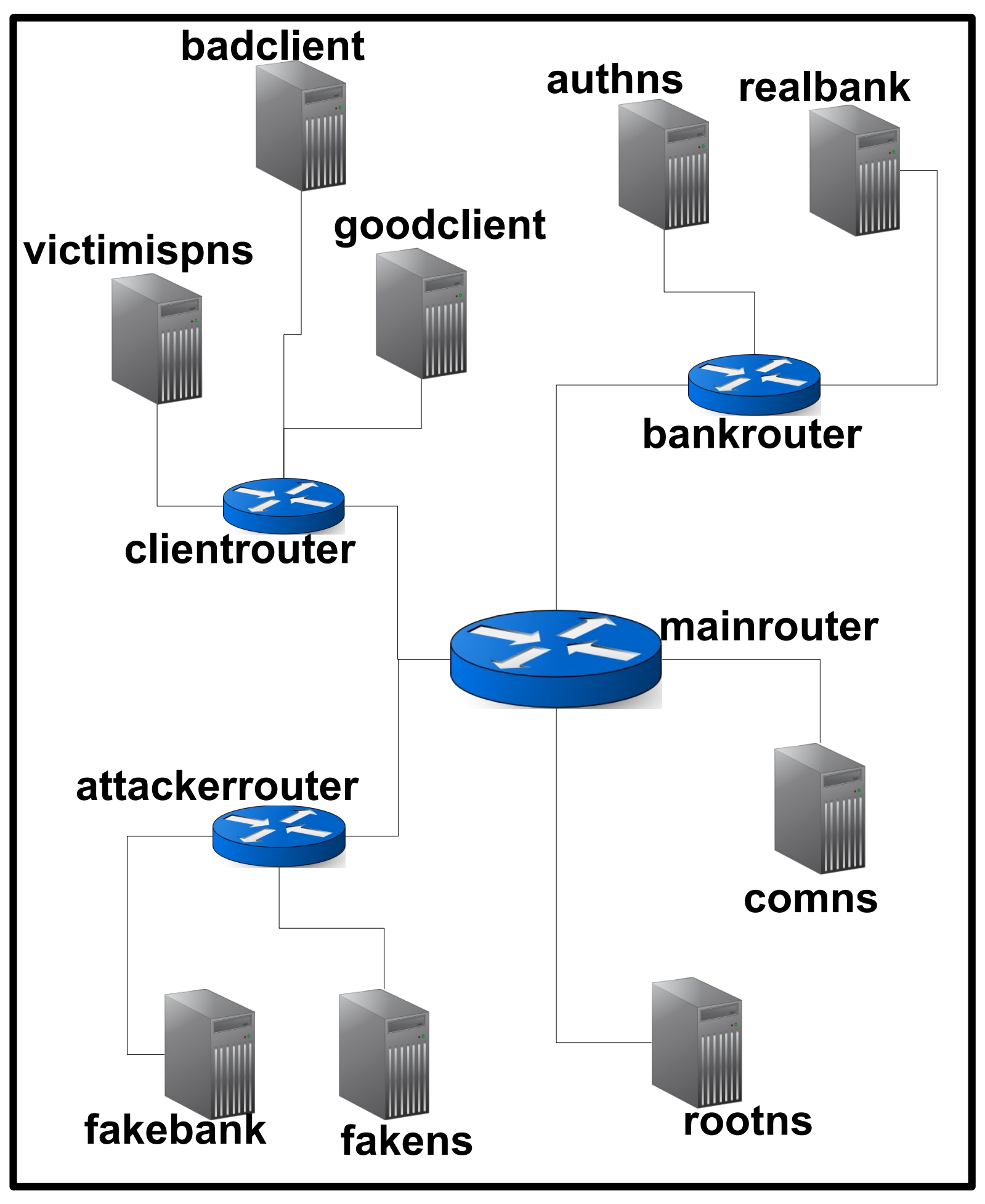


Figure 3 – Experiment topology.

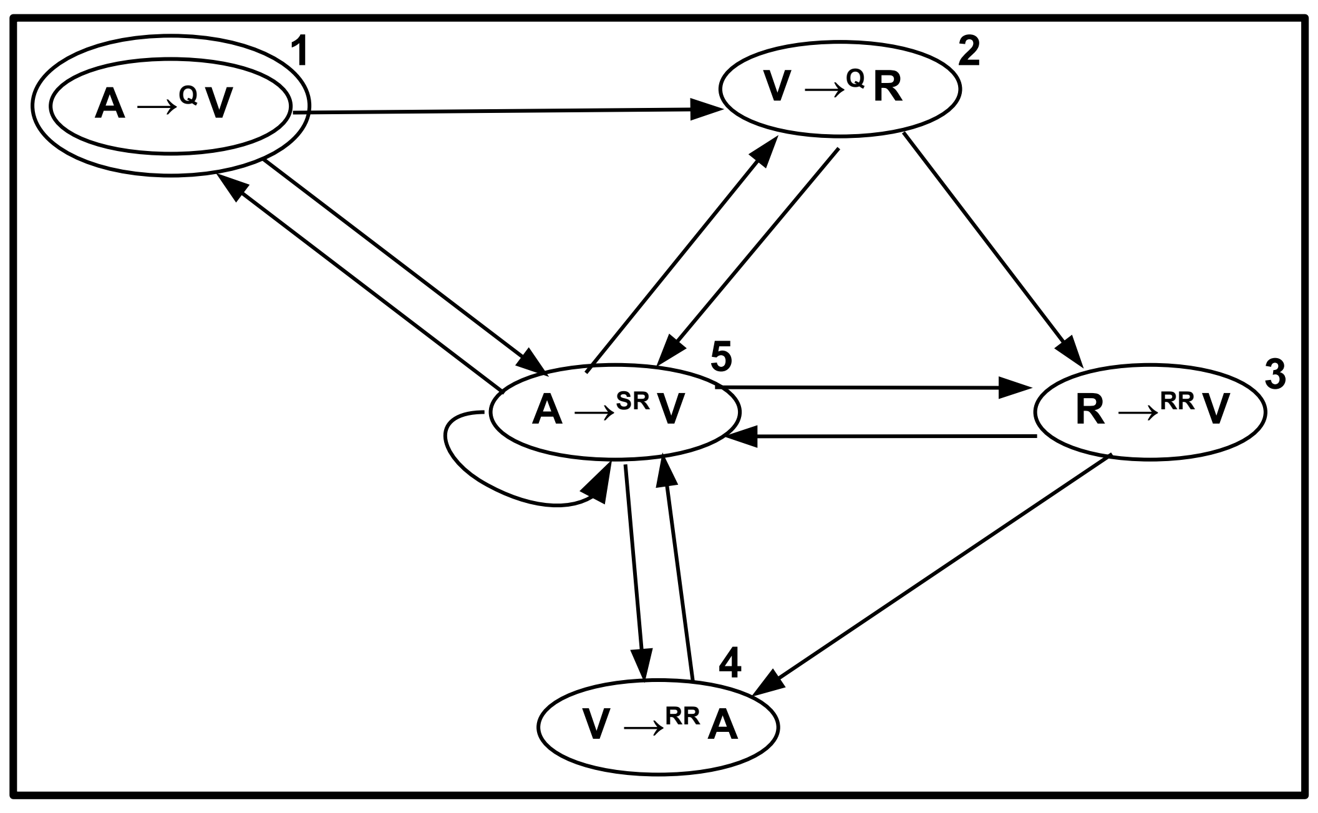


Figure 4 – Transition Diagram for sample data.

$$M(xGN)=(S, IS, A). \qquad (5)$$

$$S=\{1,2,3,4,5\}. \qquad (6)$$

$$IS=\{1\}. \qquad (7)$$

$$A=\begin{bmatrix} 0.0499 & 0.4705 & 0 & 0 & 0.4753 \\ 0.0397 & 0 & 0.4880 & 0 & 0.4688 \\ 0 & 0 & 0 & 0.3942 & 0.6051 \\ 0.0647 & 0 & 0 & 0 & 0.9344 \\ 0.0872 & 0.0518 & 0.0497 & 0.0598 & 0.7503 \end{bmatrix} \qquad (8)$$

Figure 5 – Model created from sample data.

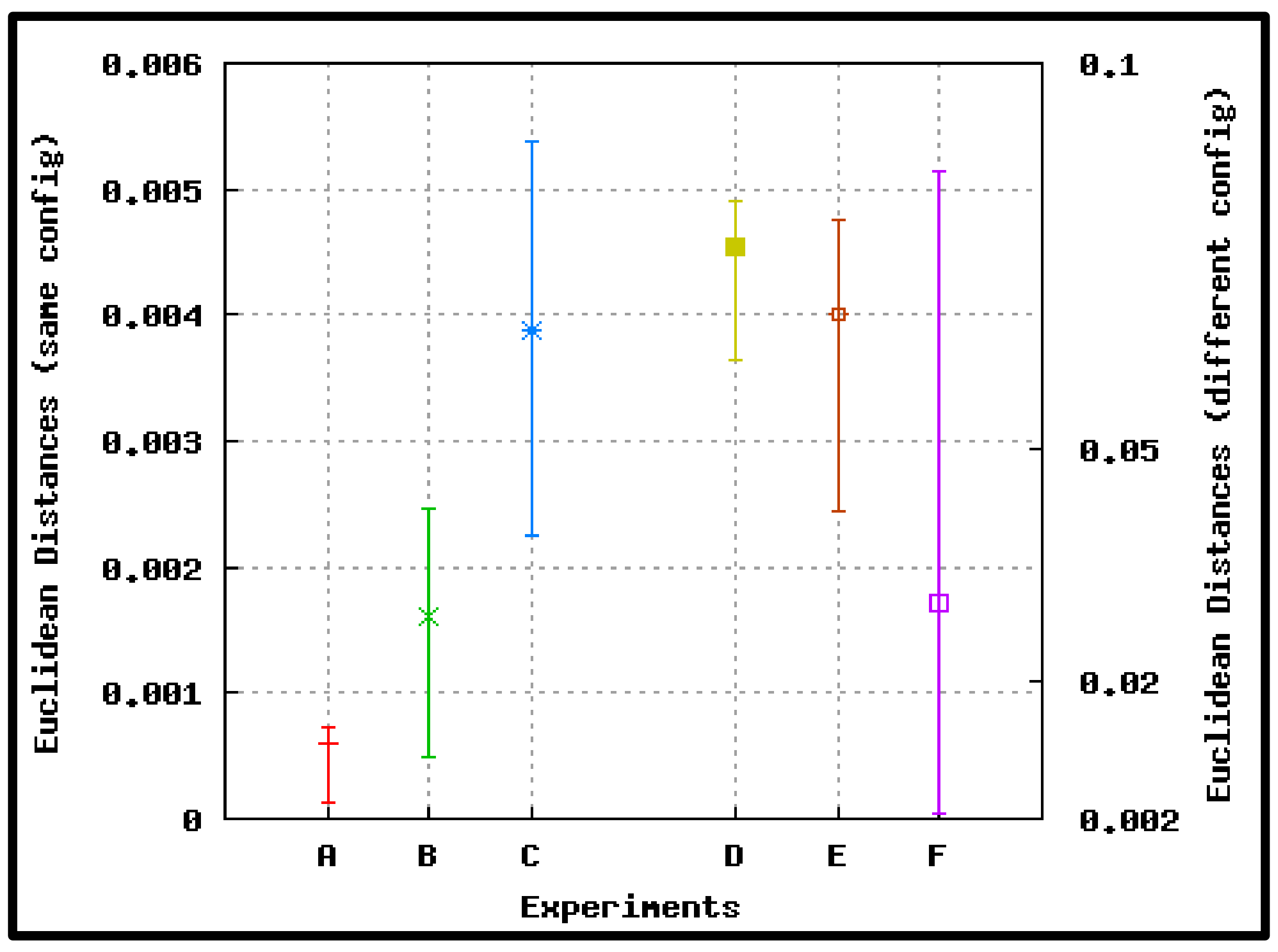


Figure 6 – Euclidean Distances for different config's.

- **Experiment Variations:**

I. Topological Variations -

(a)'authNS'- same subnet as 'victimNS'.

(b)'authNS'- same subnet as 'realbank' (global w.r.t. 'victimNS' subnet').

II. Cross-Traffic Variations -

(a) No Background Traffic.

(b) Additional DNS Traffic.

A – I (a) and II (a) comparison.

B – I (b) and II (a) comparison.

C – I (b) and II (b) comparison.

D – A and C comparison.

E – B and C comparison.

F – A and B comparison.

- **Error Bars** – Min, Median, Max – lowest to highest value .

Results

Promising methodology:

- Comparison with same config. → negligible δ (i.e. A/B/C).
- Comparison with different config. → high δ (~ 0.08 for D/E/F).
- So, small δ → same expt runs; large δ → changes in expt. config. or comparison with different expt.

Future Work

- Comparing expts in simulations, real environments to cover all kinds of experimental methodologies.
- k-order MM/HMMs → complex expts.