Voting Multi-Dimensional Data with Deviations for Web Services under Group Testing

W.T. Tsai, Yinong Chen, Dawei Zhang, Hai Huang
Software Research Laboratory
Computer Science & Engineering Department
Arizona State University
U.S.A
http://ASUSRL.EAS.ASU.EDU/SRLAB/

Service-Oriented Architecture (SOA) and Web Services (WS)

Software industry: from product-oriented model to service-oriented model

- Sign up a service contract → computer and hardware could be free
- Computers are used as a service terminal: Little maintenance

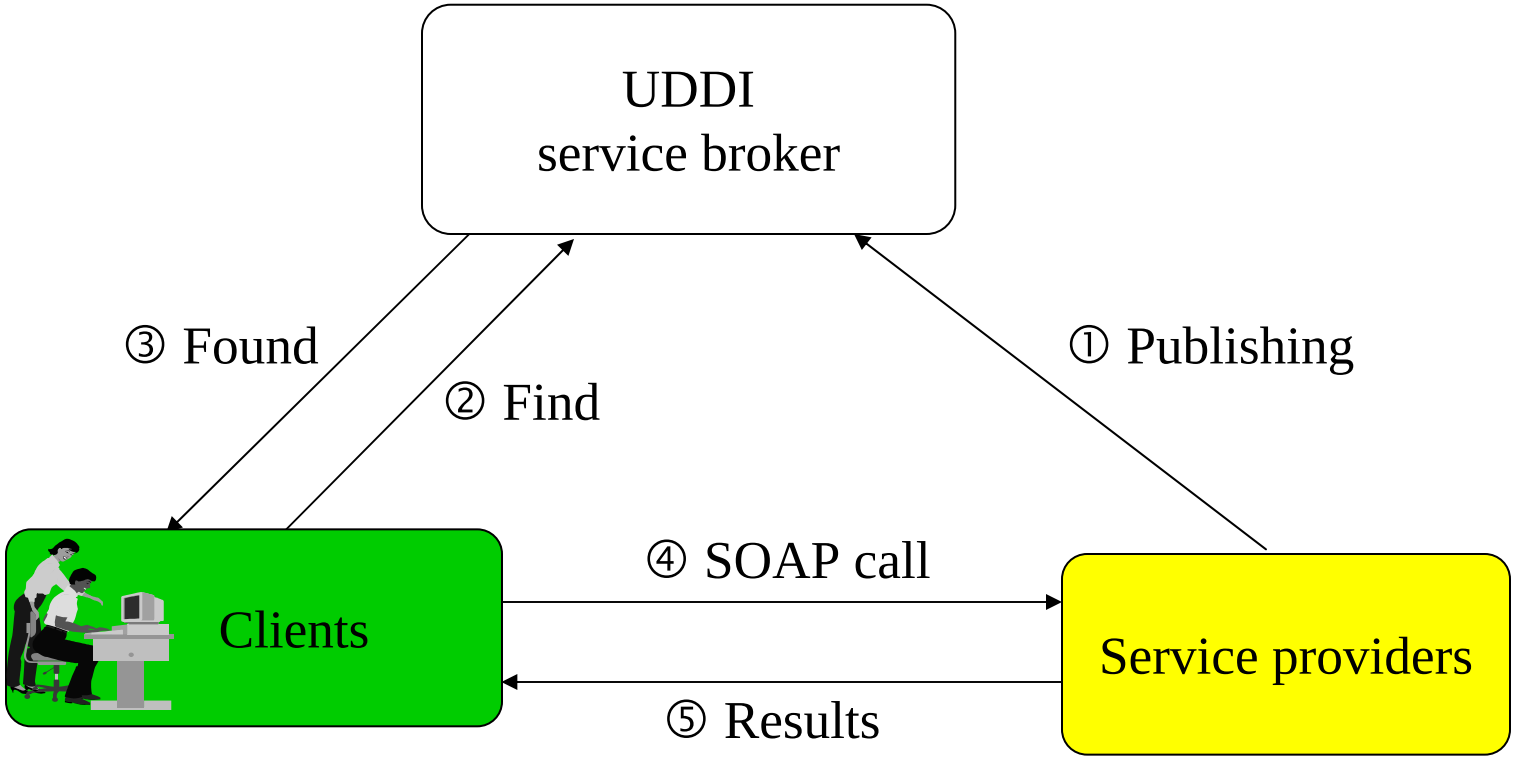
Industrial Applications

- e-Business: Security
- Mission-critical systems: Reliability
- Companies such as MS, IBM, and Sun are collaborating on application of WS, and MS is pushing WS to embedded systems.

Development of a global infrastructure to support service-oriented architecture

- Database of Web services, with ranking of trustworthiness
- Database of test scripts, their oracles, and ranking
- Reliability database

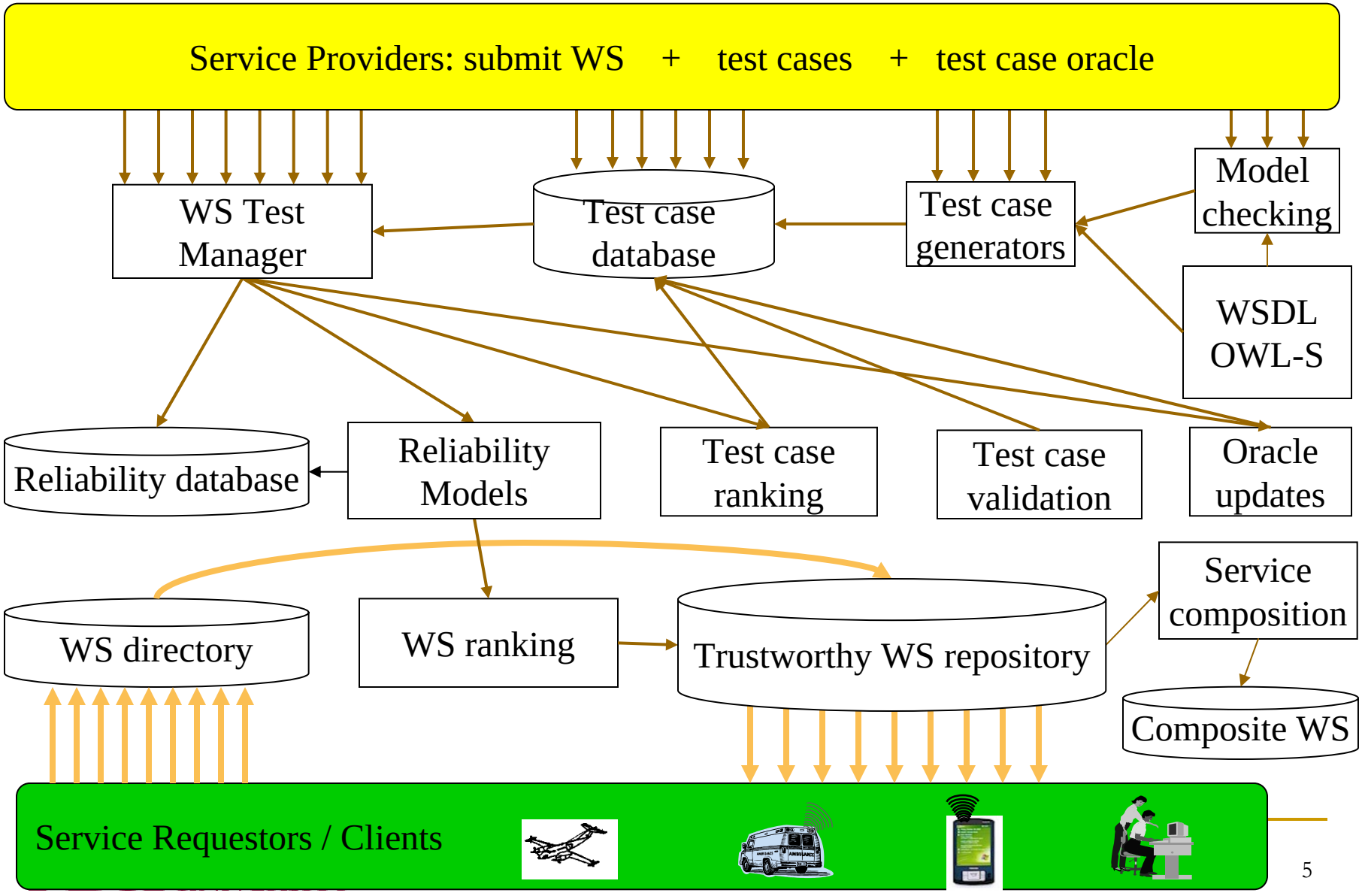
Current Web Services Model



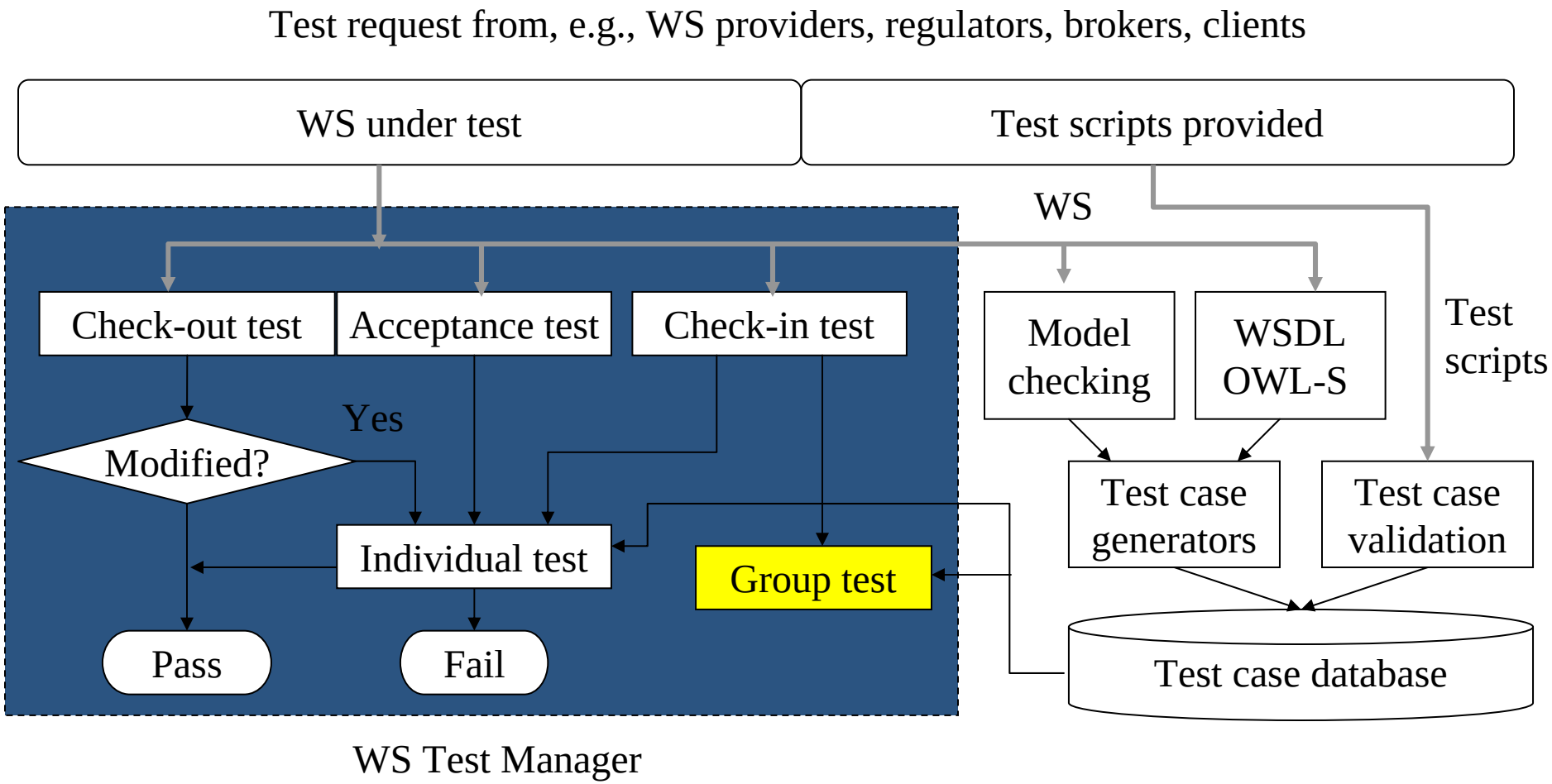
Testing of Web Services (WS)

- Need to focus on specification-based testing because WS providers may not be willing to share their source code/design.
- Need to test and evaluate a large number of WS efficiently, at runtime, and in real time.
- While test inputs can be quickly generated from WS specification (several techniques are already available to do so). It is more difficult to generate the **oracles** for these test inputs.
- WS group testing addresses this problem by using a majority voting to establish the oracles based on the premise that incorrect WS likely produce different incorrect outputs and thus oracles can be established statistically if sufficient number of WS are available.
- Each participating WS must implement the same WS specification, and thus theoretically they should produce the same or close answers, given the same inputs.

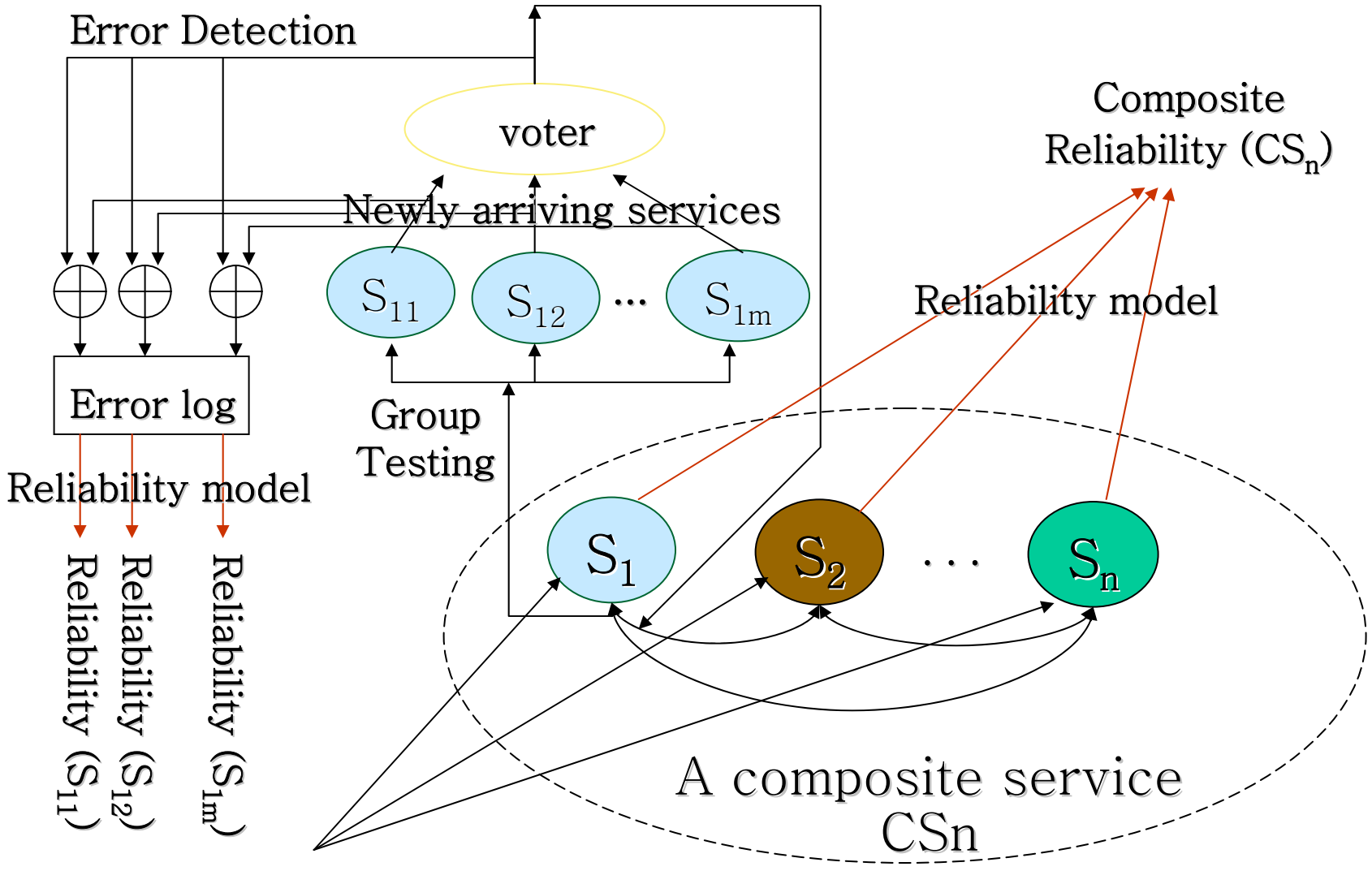
WebStrar Infrastructure @ ASU



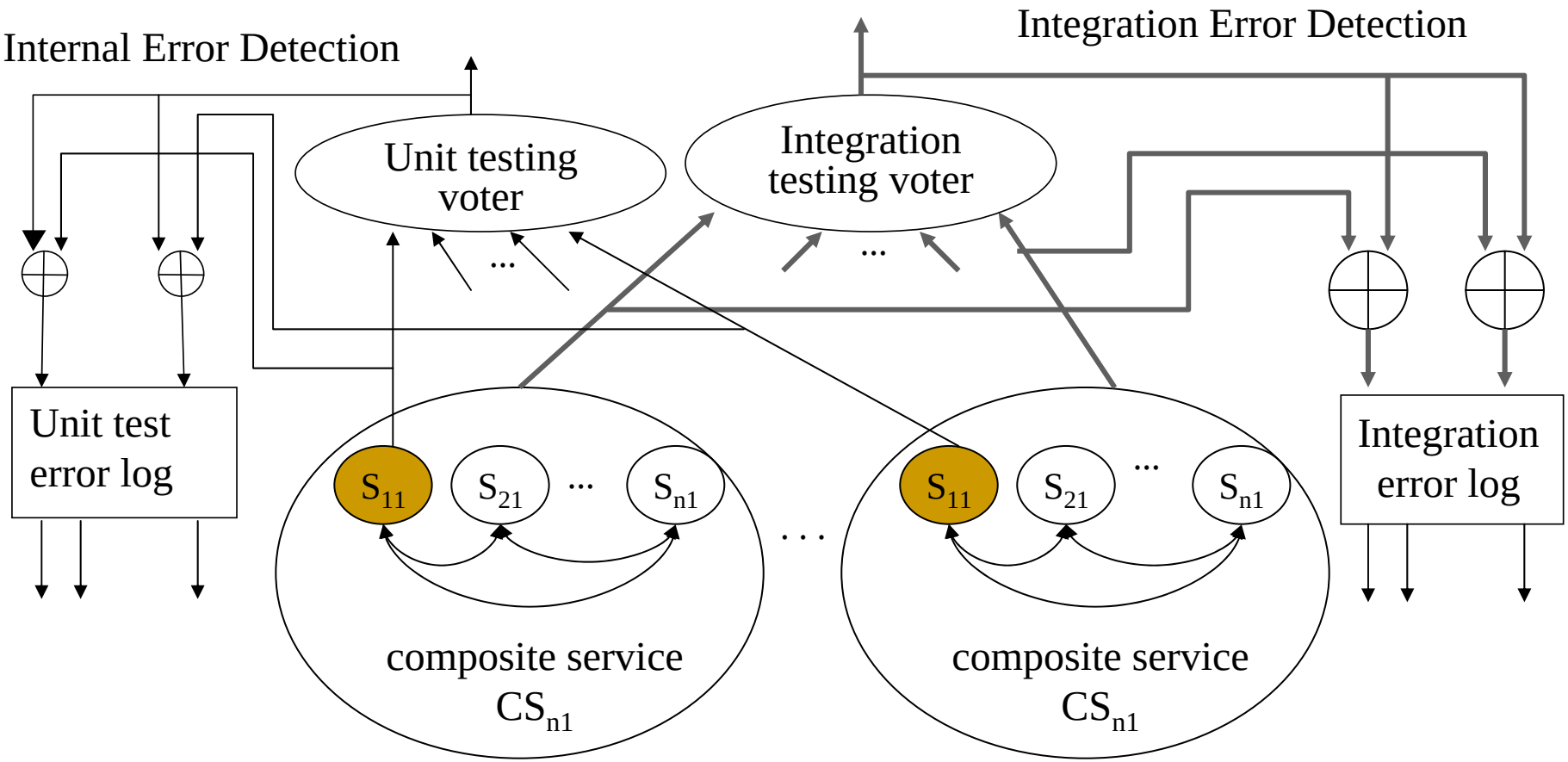
Test Manager in WebStrar Infrastructure



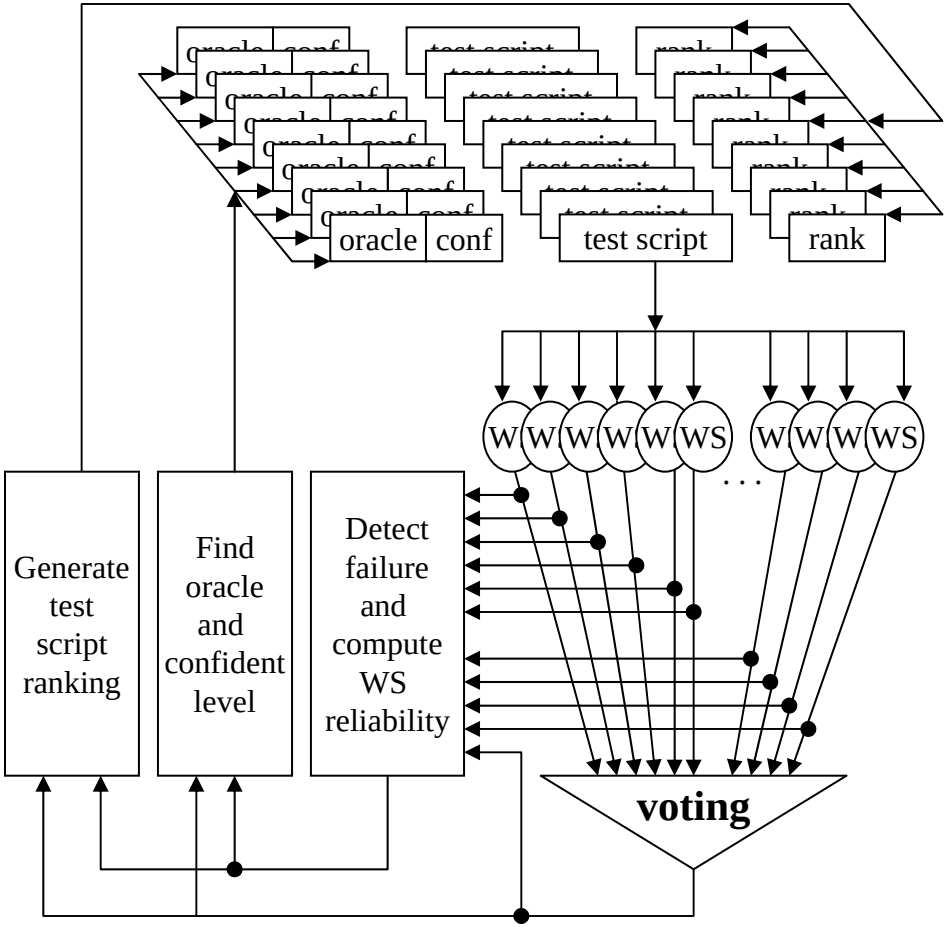
Unit Level WS Group Testing in WebStrar



Integration Level WS Group Testing



WS Group Testing and Evaluation



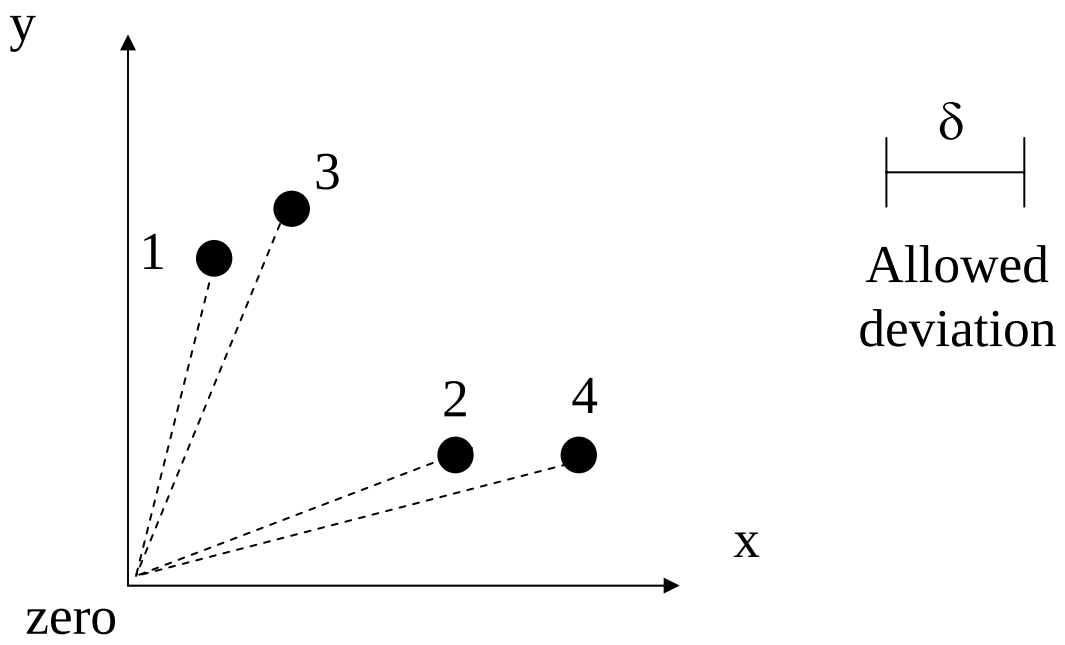
Algorithm: Clustering multi-dimensional objects

The Objects to be voted represent points in a multi-dimensional space. The data in the same cluster are those points that are close to each other.

1. Input O = {O_1, O_2, …., O_m};
2. SetClusters = { };
3. If O is empty, output ClusterSet, exit.
4. Randomly choose one object *c* to start a new cluster C = {*c*};
5. Compute the geometric center point *u* of the objects in C;
6. Computer the distance $d(u, v)$, where, $v \in O$;
7. Find the object *x* with minimum distance to *u*: $d(u, x) \leq d(u, v)$;
8. If $d(u, x) \leq \delta$ (the allowed deviation) then move *x* from O to C: $C = C \cup \{x\}$ and $O - \{x\}$;
9. Else a cluster is found: ClusterSet = ClusterSet $\cup$ {C}, goto step 3.

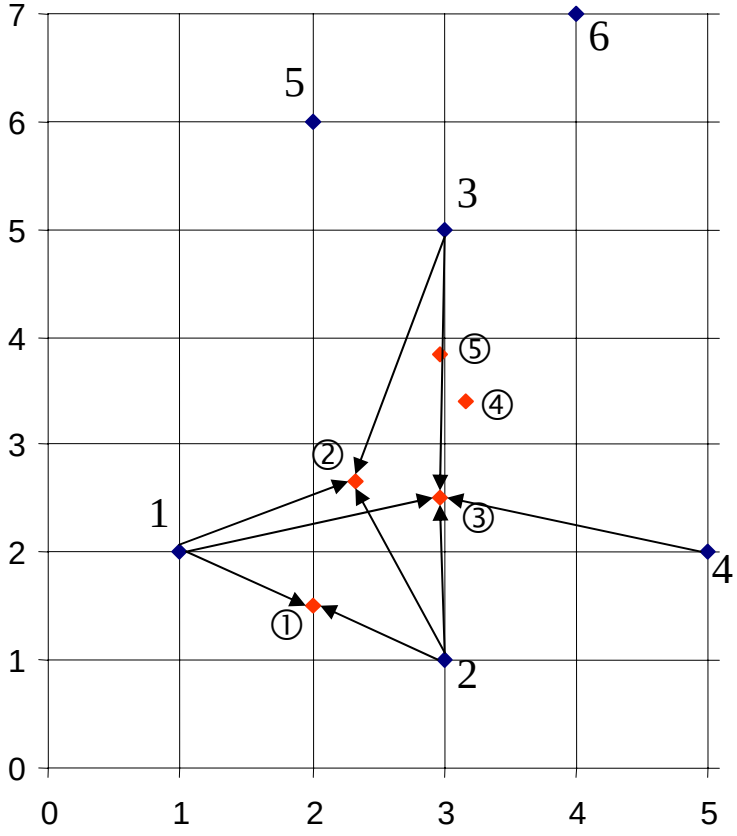
Time complexity: $O(k*m^2)$, k is dimension and m is the number of objects

Using Relative Distance vs. Absolute Distance



Finding the Geometric Central Point in a Cluster

Point number	Point (x y)	Central point	Central point number
1	(1 2)	(1.0 2.0)	
2	(3 1)	(2.0 1.5)	①
3	(3 5)	(2.3 2.7)	②
4	(5 2)	(3.0 2.5)	③
5	(2 6)	(2.8 3.2)	④
6	(4 7)	(3.0 3.8)	⑤

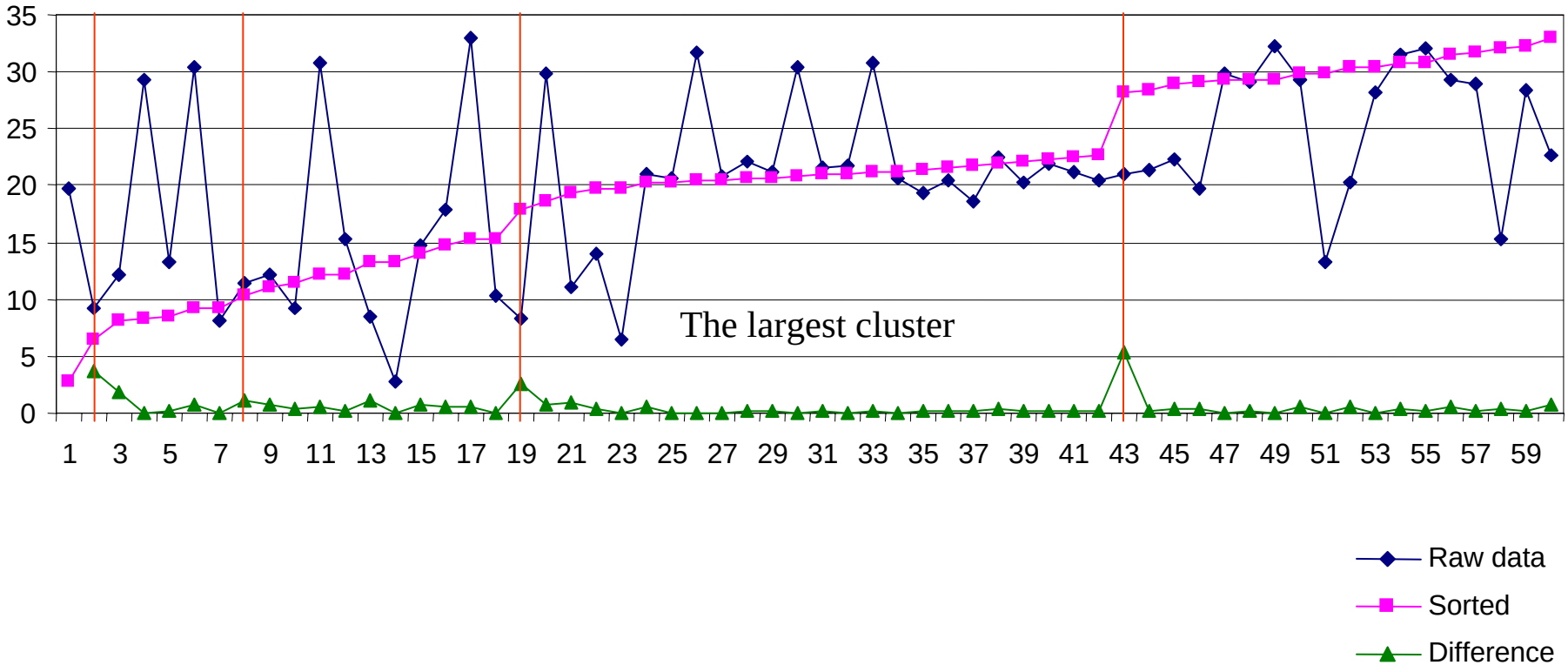


Data Collected for Voting

Table 2. Experiment data: stock prices from 60 WS, oracle, and the best 14 WS

Stock Price	Stock Price	Stock Price	Stock Price	Oracle = 20.81	
				Rank	Best Value
19.8	10.36	21.5	19.8		
9.15	8.24	21.71	29.8	1	20.77
12.24	9.80	30.73	29.15	2	20.72
29.3	11.04	**20.62**	32.24	3	20.93
13.31	14.03	19.41	29.3	4	20.93
30.35	6.41	**20.47**	13.31	5	20.62
8.17	20.93	18.58	**20.35**	6	21.12
11.49	**20.72**	22.56	28.17	7	21.12
12.13	21.75	**20.28**	31.49	8	20.47
9.2	**20.77**	21.99	32.13	9	20.41
30.85	22.10	**21.12**	29.2	10	20.35
15.22	**21.12**	**20.41**	28.85	11	20.28
8.41	**20.39**	**20.93**	15.22	12	21.35
2.75	**21.50**	**21.35**	28.41	13	21.50
14.67	**21.71**	22.34	22.75	14	21.71

Clustering of the Data



Enhanced Stochastic Voting

1. Convert non numerical results into numerical results and represents each WS output as an n-element object of numerical values;
2. Compute the distance matrix between each pair of objects and find the two objects with the minimum distance;
3. Apply the **Simulated Annealing** algorithm to group the objects with the same property into one cluster;
4. Applying the **Chi-Square Goodness-Fit-Test** to determine if the current cluster contains the objects that present the correct output of the WS under test;
5. If Chi-Square Goodness-Fit-Test fails, return to step 3; otherwise, terminate.

Conclusions http://ASUSRL.EAS.ASU.EDU/SRLAB/

- SOA and WS represent a new computing paradigm
- Large number of available WS online
- Specification-based testing
- Group testing technique to efficiently and effectively test and evaluate a large number of WS
- Establish the oracle statistically
- Voting of
  - numerical data,
  - data with deviations, and
  - multi-dimensional data