Optimised Realistic Test Input Generation

Mustafa Bozkurt and Mark Harman

CREST Centre,

University College London

Optimised

What is Realistic Test Input ?

Generation

Realistic data:

Structural realism: Must conform to syntactic constraints.

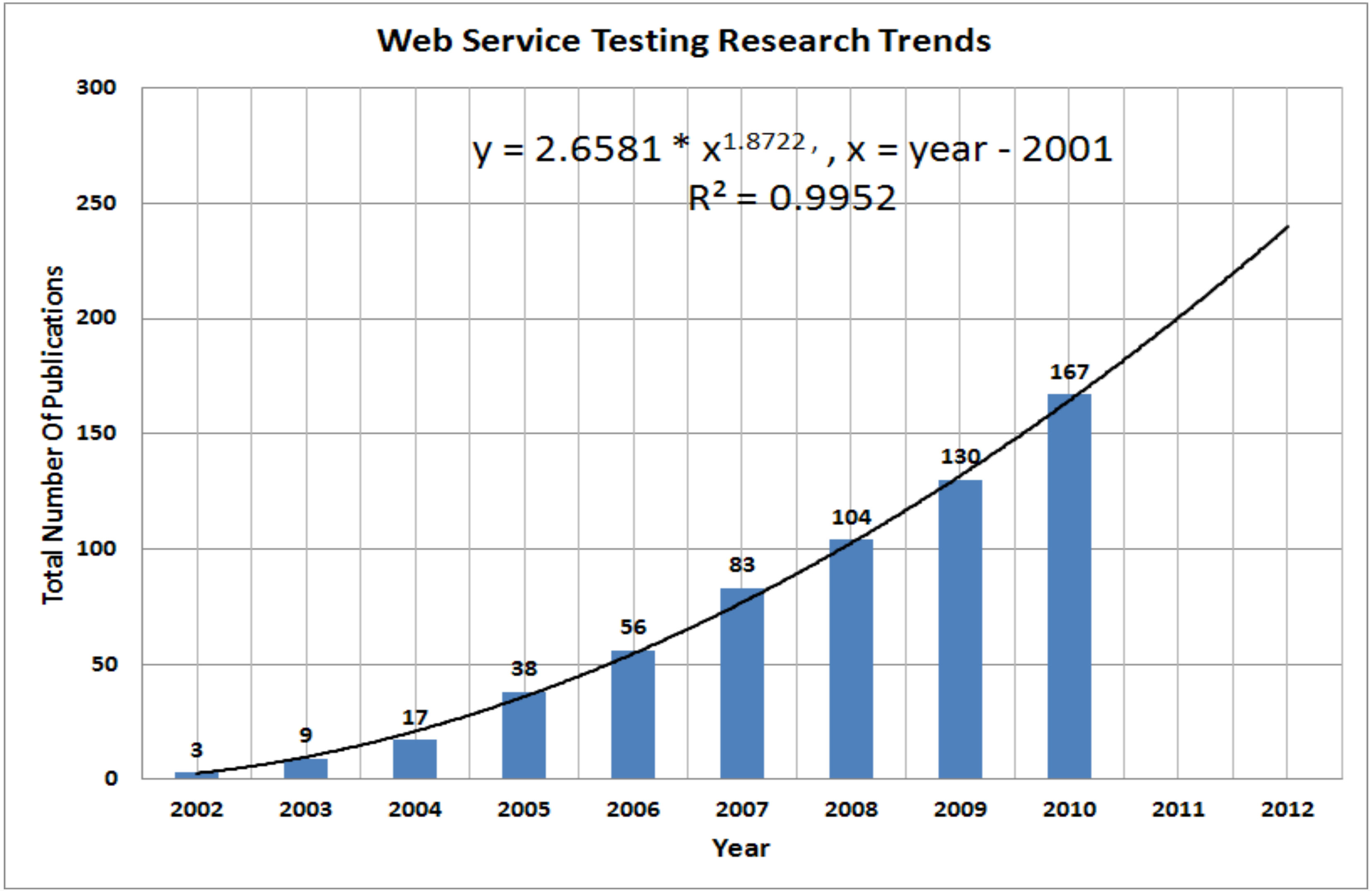
ISBN 10 digit $x_1, \ldots, x_{10}$, such that:

$$x_{10} = 11 - (10x_1 + 9x_2 + \ldots + 2x_9) \bmod 10$$

Syntactic realism: Must represent a real-world entity.

ISBN must represent a book.

WHY is realistic test data important?



- According to Microsoft (http://msdn.microsoft.com/en-us/library/aa833244.aspx)

```
private string _state;  private int _zipCode; private string _address;


private static readonly List<String> _states =
        new List<string>() { "AK", "AL", "AK", "AZ", "AR", "CA", "CO", "CT", "DE", "FL",
"GA", "HI", "ID", "IL", "IN", "IA", "KS", "KY", "LA", "ME", "MD", "MA", "MI", "MN",
"MS", "MO", "MT", "NE", "NV", "NH", "NJ", "NM", "NY", "NC", "ND", "OH", "OK", "OR",
"PA", "RI", "SC", "SD", "TN", "TX", "UT", "VT", "VA", "WA", "WV", "WI", "WY" };

protected override void OnGenerateNextValues(){
   this._street = "2150 Newton Street";
        this._city = "San Francisco";
        this._state = _states[Random.Next(0, _states.Count)];
        this._zipCode = Int32.Parse(String.Format("{0}{1}{2}{3}{4}", GetRandomDigit(1),
GetRandomDigit(), GetRandomDigit(), GetRandomDigit(), GetRandomDigit()));
        this._address = String.Format("{0}, {1}, {2} {3}", _street, _city, _state, _zipCode);
```

??? Holtsville, NY 00501

- Mcminn et al.

P. McMinn, M. Stevenson, and M. Harman, "Reducing qualitative human oracle costs associated with automatically generated test data," in Proceedings of the 1st International Workshop on Software Test Output Validation (STOV 2010). Trento, Italy: ACM Press., July 2010, pp. 1–4.

Automated, provides only structural realism

- Konroy et al.

K. Conroy, M. Grechanik, M. Hellige, E. Liongosari, and Q. Xie, "Automatic test generation fromGUI applications for testing web services," in ICSM 2007: Proceedings of the 23rd IEEE Interna-tional Conference on Software Maintenance. Paris, France: IEEE Computer Society, October 2007,pp. 345–354.

Hard to automate, realistic

- Alshahwan and Harman

N. Alshahwan and M. Harman, "Automated session data repair for web application regression test-ing." in ICST'08: Proceedings of the 2008 International Conference on Software Testing, Verification,and Validation. Lillehammer, Norway: IEEE Computer Society, April 2008, pp. 298–307.

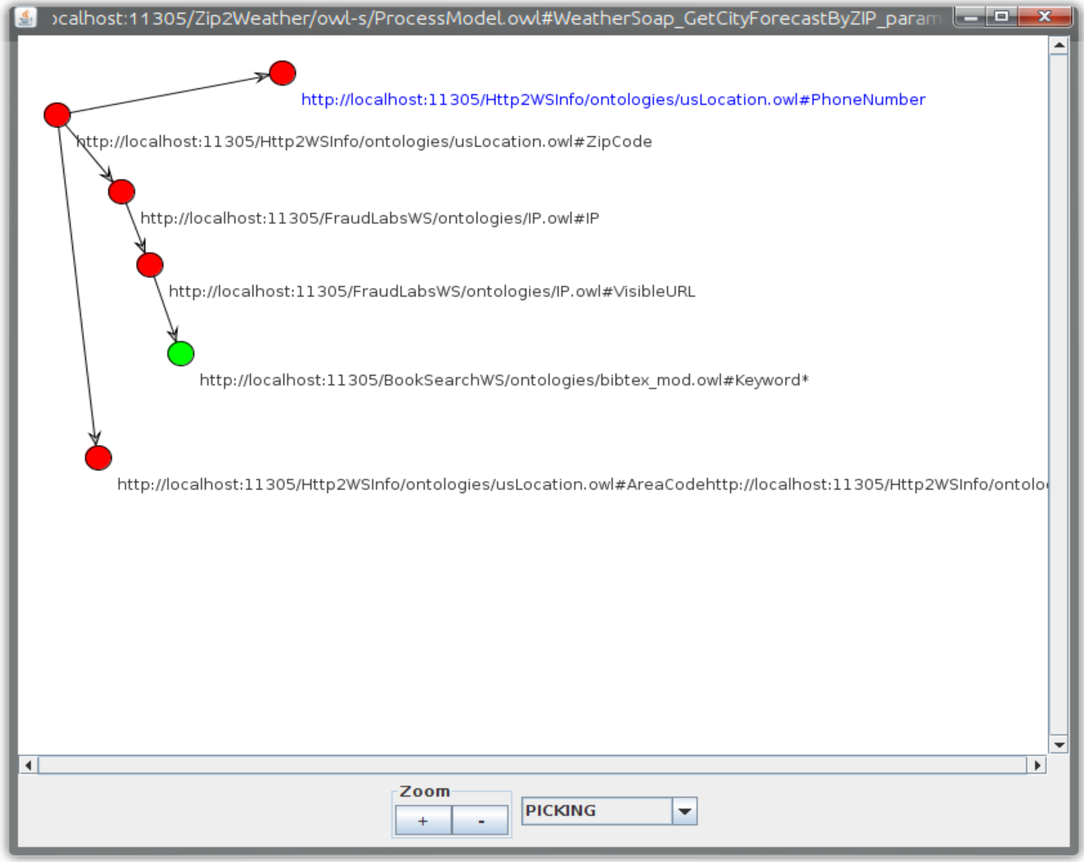
Automated, realistic but highly dependent on existing resources

A novel automated solution to realistic test data generation that exploits existing web services as sources of realistic test data.

Our approach is capable of:

- Generating realistic test data.
- Generating data based on tester-specified constraints.

ATAM realistic input generator


›calhost:11305/Zip2Weather/owl-s/ProcessModel.owl#WeatherSoap_GetCityForecastByZIP_param
http://localhost:11305/Http2WSInfo/ontologies/usLocation.owl#PhoneNumber
http://localhost:11305/Http2WSInfo/ontologies/usLocation.owl#ZipCode
http://localhost:11305/FraudLabsWS/ontologies/IP.owl#IP
http://localhost:11305/FraudLabsWS/ontologies/IP.owl#VisibleURL
http://localhost:11305/BookSearchWS/ontologies/bibtex_mod.owl#Keyword*
http://localhost:11305/Http2WSInfo/ontologies/usLocation.owl#AreaCodehttp://localhost:11305/Http2WSInfo/ontolo
Zoom
+
-
PICKING

ZIP code generation case study

- 31 search engines
- 13 WS to IP
- 40 IP to Location

31x13x40 = 16,120 possible solutions

7-service sequence with 15 alternatives $15^7 \approx 171\text{x}10^6$

Bozkurt, M., Harman, M.: Generating realistic test input using semantic web services. Tech. Rep. RN/11/17, University College London TO APPEAR SOSE 2011

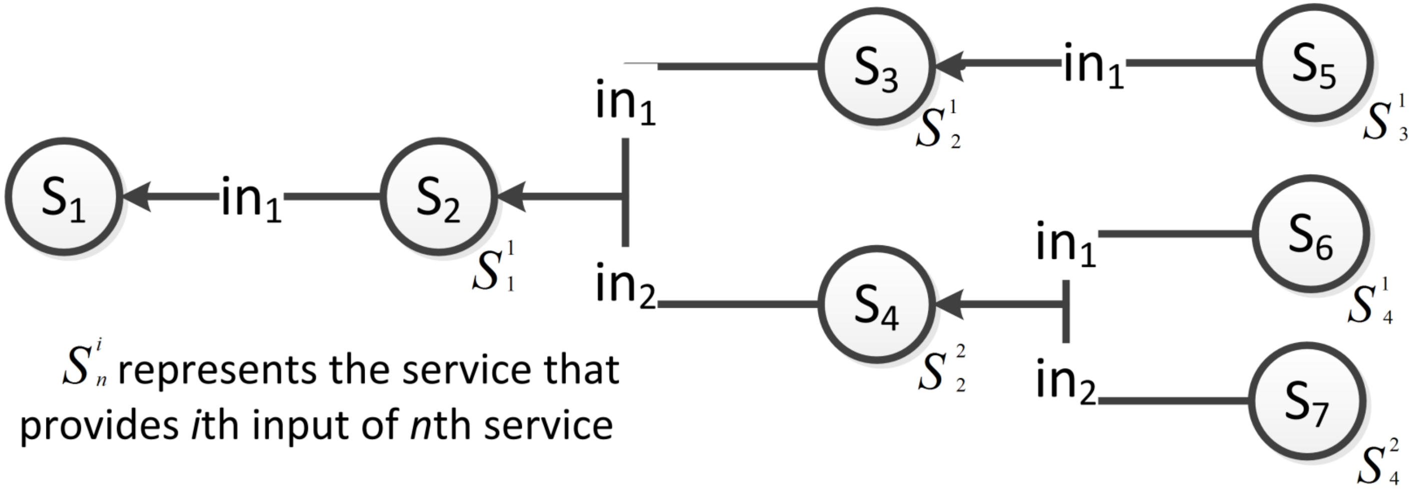
ATAM needs to:

1. Select and use services with high reliability.
2. Select and use services with low execution price.

Advantage of semantic SOA:

1. Ability to discover services automatically.
2. Ability to invoke services automatically.
3. QoS parameters which include: reliability, execution price, performance, availability and security*

*Wan Ab. Rahman, W., Meziane, F.: Challenges to describe QoS requirements for web services quality prediction to support web services interoperability in electronic commerce. In: Proceedings of the 10th IBIMA Conference on Innovation and Knowledge Management in Business. vol. 4, pp. 50-58. International Business Information Management Association (IBIMA), Kuala Lumpur, Malaysia (June 2008)

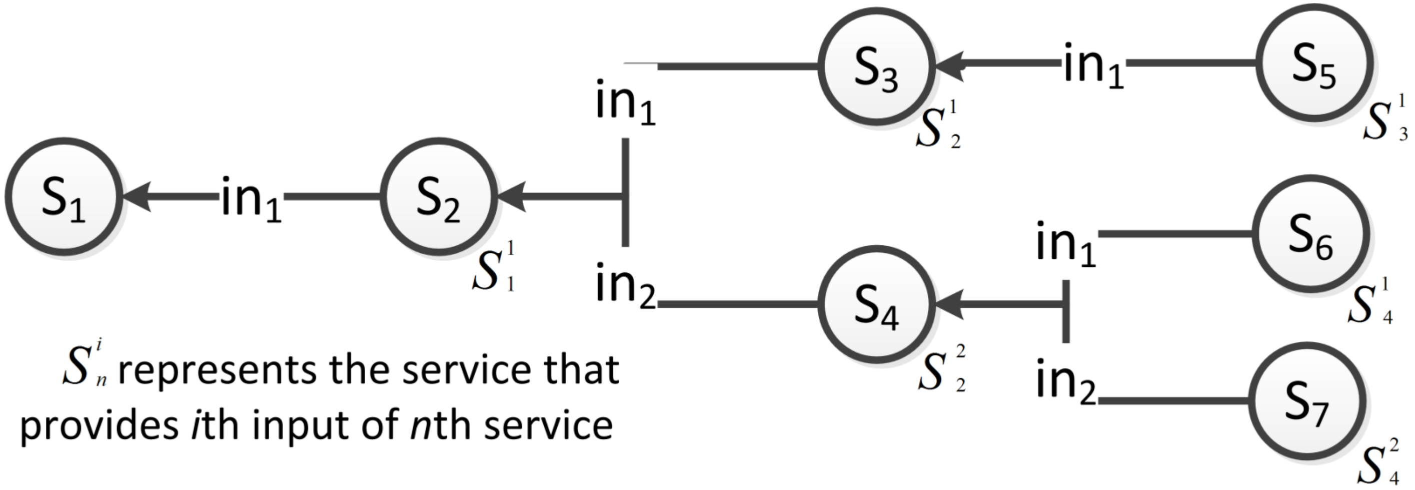


S_n^i represents the service that provides *i*th input of *n*th service

Objective Functions:

1. Cost:

$$Minimise \sum_{i=1}^{n} P_{S_i}$$

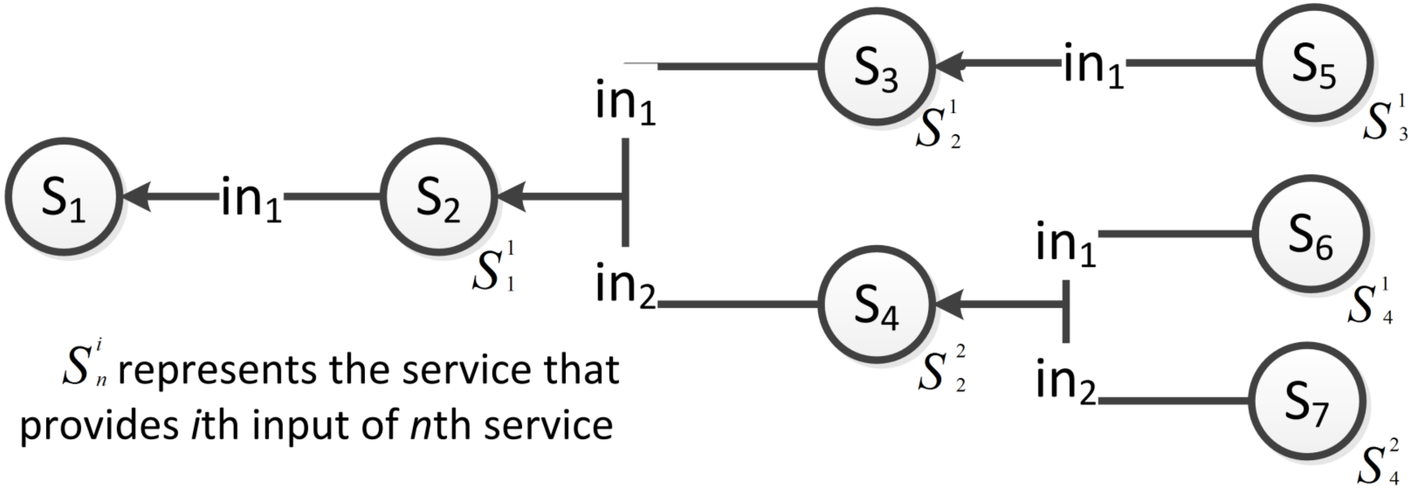


S_n^i represents the service that provides *i*th input of *n*th service

Objective Functions:

2. Reliability:

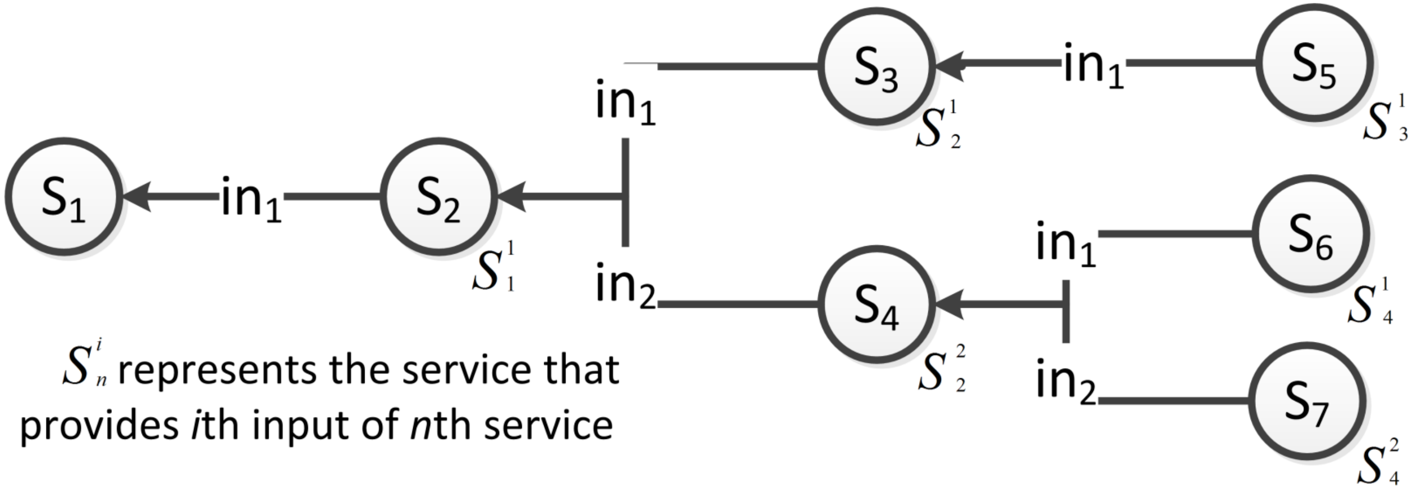
$$cr(S_n) = r_{S_n}\ x\ ir(S_n)$$

- 

S_n^i represents the service that provides *i*th input of *n*th service

2. Reliability:

$$ir(S_n) = \begin{cases} 1.0 & \text{if } S_n \text{ uses tester/generated input} \\ \mathrm{Min}(cr(S_n^1), cr(S_n^2), \ldots, cr(S_n^{in(S_n)})) & \text{if } S_n \text{ uses service output} \end{cases}$$



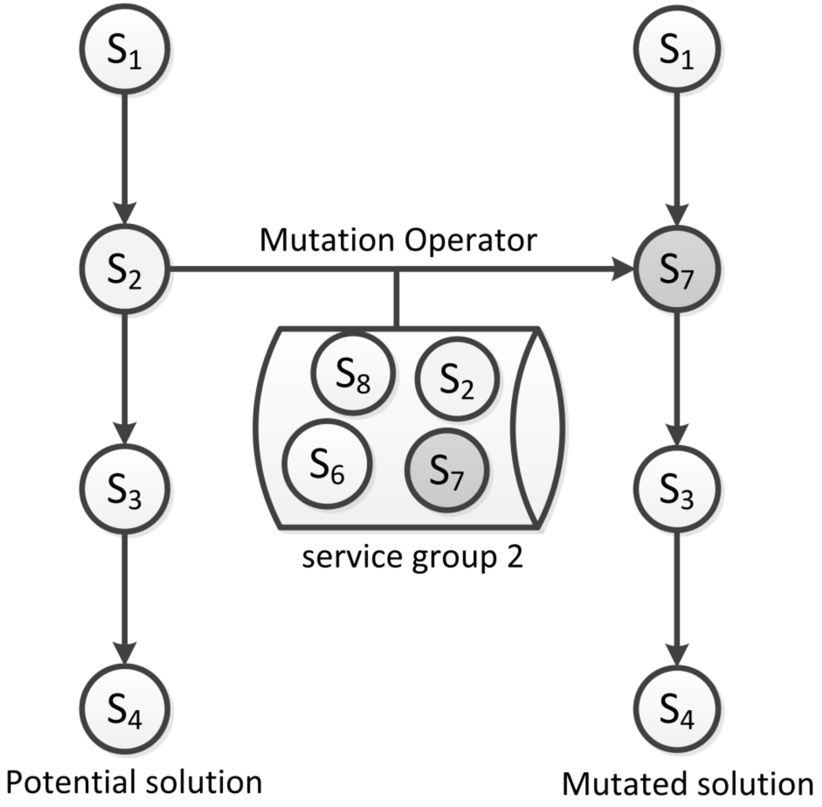
S_n^i represents the service that provides *i*th input of *n*th service

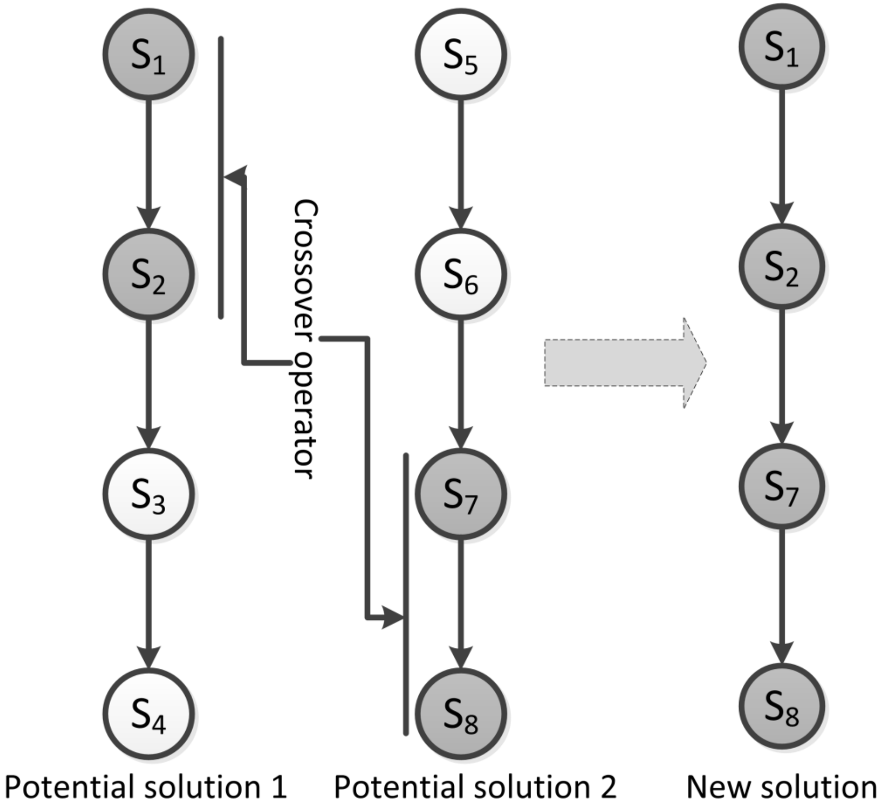
Objective Functions:

2. Reliability:

$$Maximise\ cr(S_1) = r_{S_1}\ x\ ir(S_1)$$

Generic Operators:





Optimised Realistic Input Generation:

1. Selection of test data source.
2. Increased data and input generation reliability.
3. Reduced test data generation cost.