

SOAP and BLAST: New network BLAST services at the NCBI



Christiam Camacho, Vahram Avagyan, George Coulouris, Thomas Madden
 National Center for Biotechnology Information, NIH, Bethesda, MD 20894
 {camacho,avagyanv,coulouri,madden}@ncbi.nlm.nih.gov

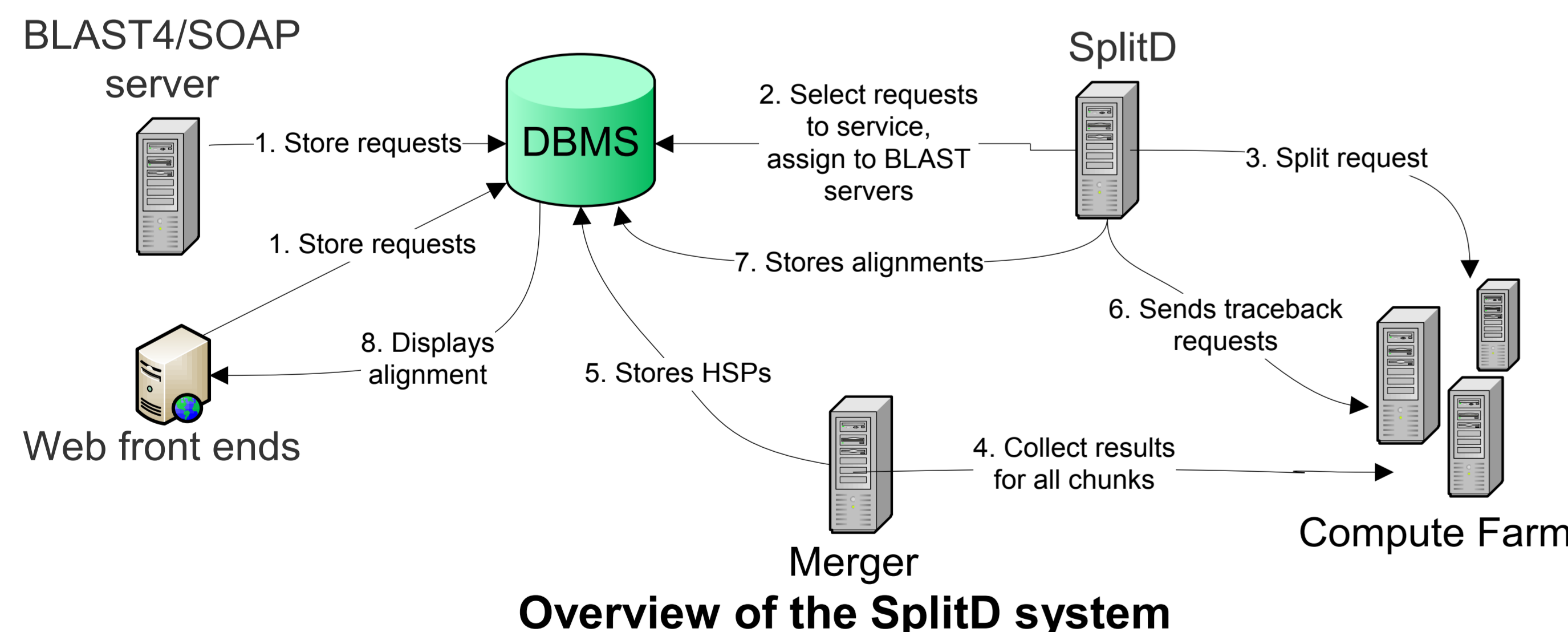


Introduction

Public access to the Basic Local Alignment Search Tool (BLAST) service offered by the National Center for Biotechnology Information (NCBI) is available through the World Wide Web as well as other network protocols. Of these we highlight a new, Simple Object Access Protocol (SOAP)-based network service which allows programmatic access to BLAST and offers many new features and improvements over previously available network services.

Today's network BLAST

- SplitDB system: introduced Q3 2003
- Served by a farm of 424 CPUs
- DBMS based queueing system
- Communication via BLAST+ command line applications (synchronous, ASN.1-encoded requests using BLAST4 network protocol) and SOAP-based web service interface
- Handles ~350k requests per day and growing



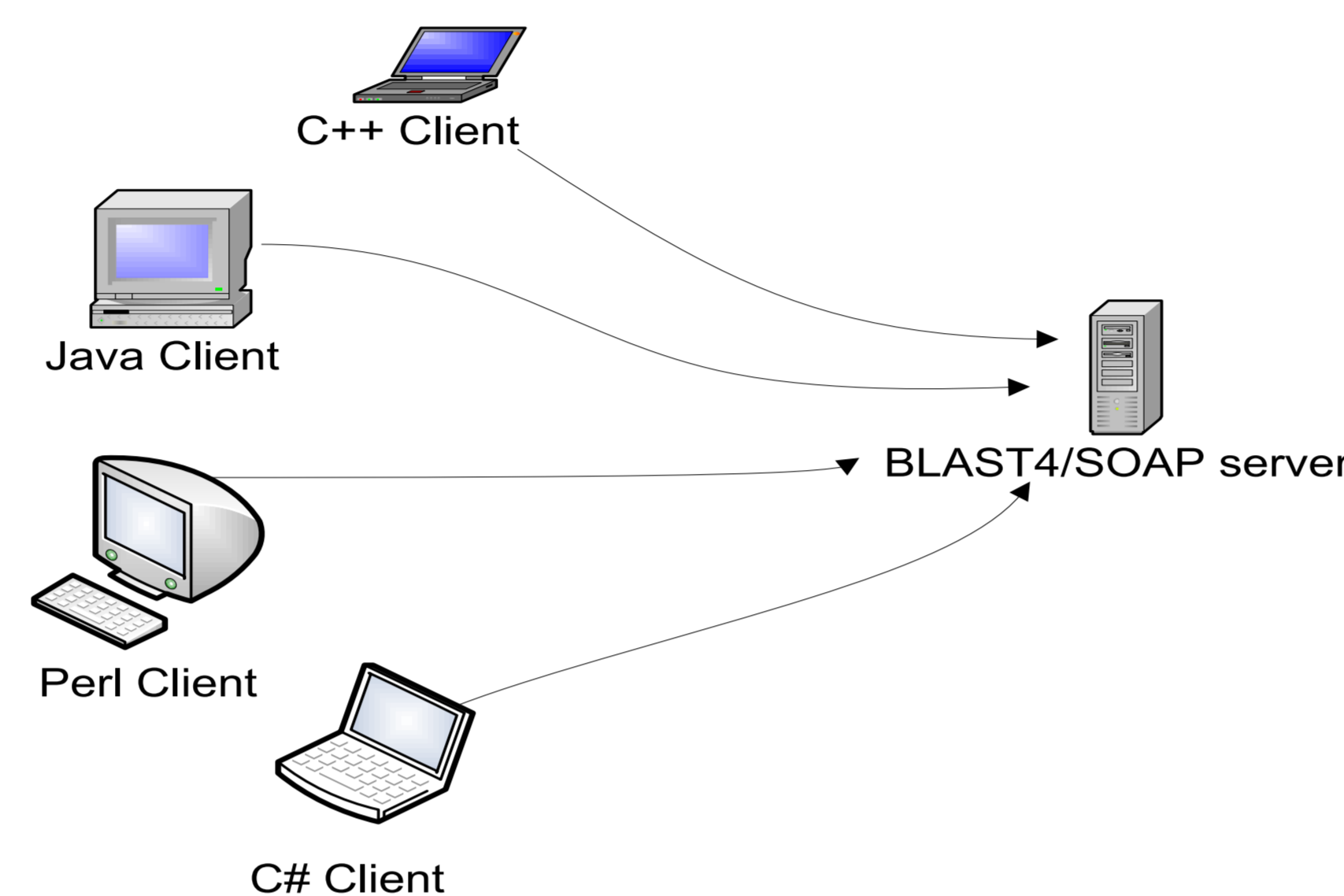
Overview of the SplitD system

Benefits of BLAST4 protocol

- Easily extensible
- Support for alignment between individual sequences (bl2seq)
- Support for new features and concepts (search strategies, tasks)

Benefits of SOAP

- Uses widely adopted industry standards (XML, WSDL)
- Interoperability



SOAP Interface offers interoperability with various programming languages

Service description

BlastWebService
+SubmitSearchLite()
+SubmitSearch()
+CheckSearchStatus()
+GetSearchResults()
+GetDatabases()
+GetSupportedMatrices()
+GetSupportedOptions()
+GetSupportedTasks()
+GetSupportedPrograms()
+GetSequences()
+GetSearchStrategy()

SOAP-based BLAST interface

- SOAP interface provides access to ample computing resources used to serve the public BLAST web pages
- Load-balanced, fault-tolerant system
- Simplified submission and results interface
- Sample clients available in C++, Java, Perl, and C#

```
<?xml version="1.0" encoding="UTF-8"?>
<soap:Envelope
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:soapenc="http://schemas.xmlsoap.org/soap/encoding/"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  soap:encodingStyle="http://schemas.xmlsoap.org/soap/encoding/"
  xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/">
  <soap:Body>
    <namesp1:Blast4-queue-search-request-lite
      xmlns:namesp1="http://www.ncbi.nlm.nih.gov">
      <query xsi:type="xsd:string">555</query>
      <database-name xsi:type="xsd:string">nt</database-name>
      <options>
        <Blast4-options-lite>
          <task xsi:type="xsd:string">blastn</task>
        </Blast4-options-lite>
      </options>
    </namesp1:Blast4-queue-search-request-lite>
  </soap:Body>
</soap:Envelope>
```

Sample SOAP Request for search submission

```
<?xml version="1.0"?>
<env:Envelope xmlns:env="http://schemas.xmlsoap.org/soap/envelope/">
  <env:Body>
    <Blast4-queue-search-reply xmlns="http://www.ncbi.nlm.nih.gov">
      <request-id>1223326474-15895-191743849031.TESTBLAST</request-id>
      </Blast4-queue-search-reply>
    </env:Body>
  </env:Envelope>
```

Sample SOAP Response for search submission

```
U:\spec.asn
-- Complete set of simple Blast results
Blast4-simple-results ::= SEQUENCE <
  all-alignments SEQUENCE OF Blast4-alignments-for-query
>
-- Alignments for one query, compiled from the raw SeqAlign results
Blast4-alignments-for-query ::= SEQUENCE <
  -- Query sequence identifier
  -- (present if query is not a local id in the SeqAlign)
  query-id VisibleString,
  -- All the alignments for this query
  alignments SEQUENCE OF Blast4-simple-alignment
>
-- A single alignment
Blast4-simple-alignment ::= SEQUENCE <
  -- Subject sequence identifier
  -- (normally a GI from the SeqAlign)
  subject-id VisibleString,
  -- E-Value
  e-value REAL,
  -- Bit score
  bit-score REAL,
  -- Number of identities
  num-identities INTEGER OPTIONAL,
  -- Number of insertions/deletions
  num-indels INTEGER OPTIONAL,
  -- Full query range covered by this HSP
  full-query-range Blast4-range,
  -- Full subject range covered by this HSP
  full-subject-range Blast4-range
>
-- Range on a sequence - zero offset -
Blast4-range ::= SEQUENCE <
  start INTEGER OPTIONAL,
  end INTEGER OPTIONAL,
  strand INTEGER OPTIONAL
>
```

Simplified BLAST results specification

- A BLAST4/SOAP server is available for testing at <http://www.ncbi.nlm.nih.gov/blast/netblast/blastws.cgi?wsdl>
- More information and demo clients are available at ftp://ftp.ncbi.nlm.nih.gov/blast/temp/web_services/

