

REMEDI: Robust and Efficient Machine Translation in a Distributed Infrastructure

Bi-Annual Report: Month 18

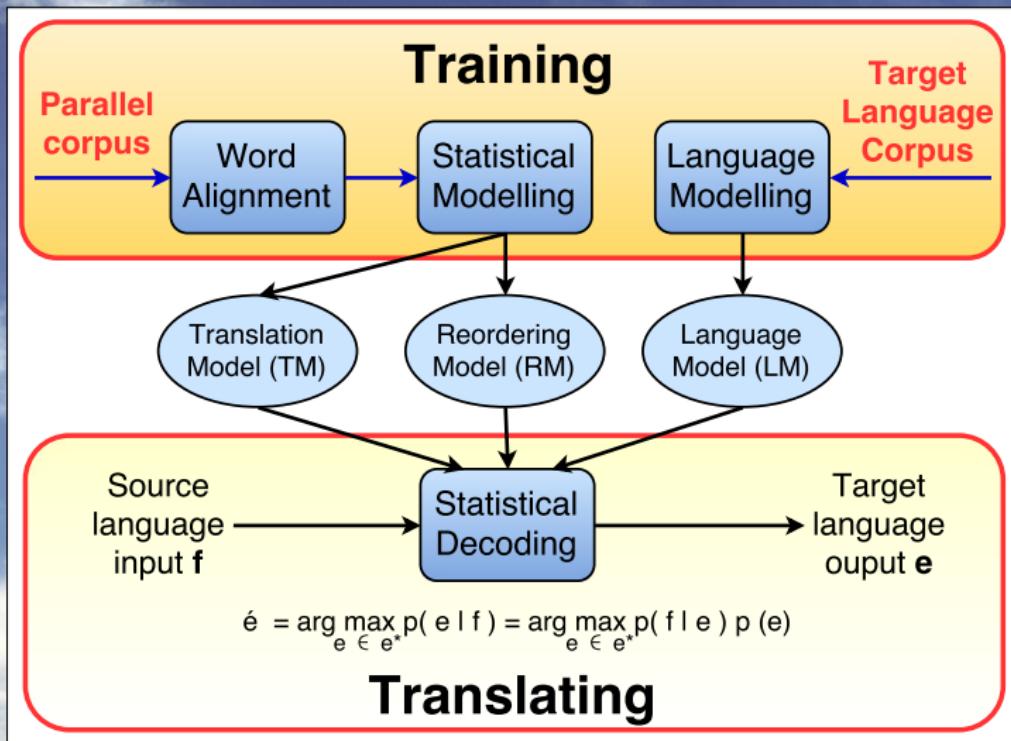
Dr. Ivan S. Zapreev

Dr. Christof Monz

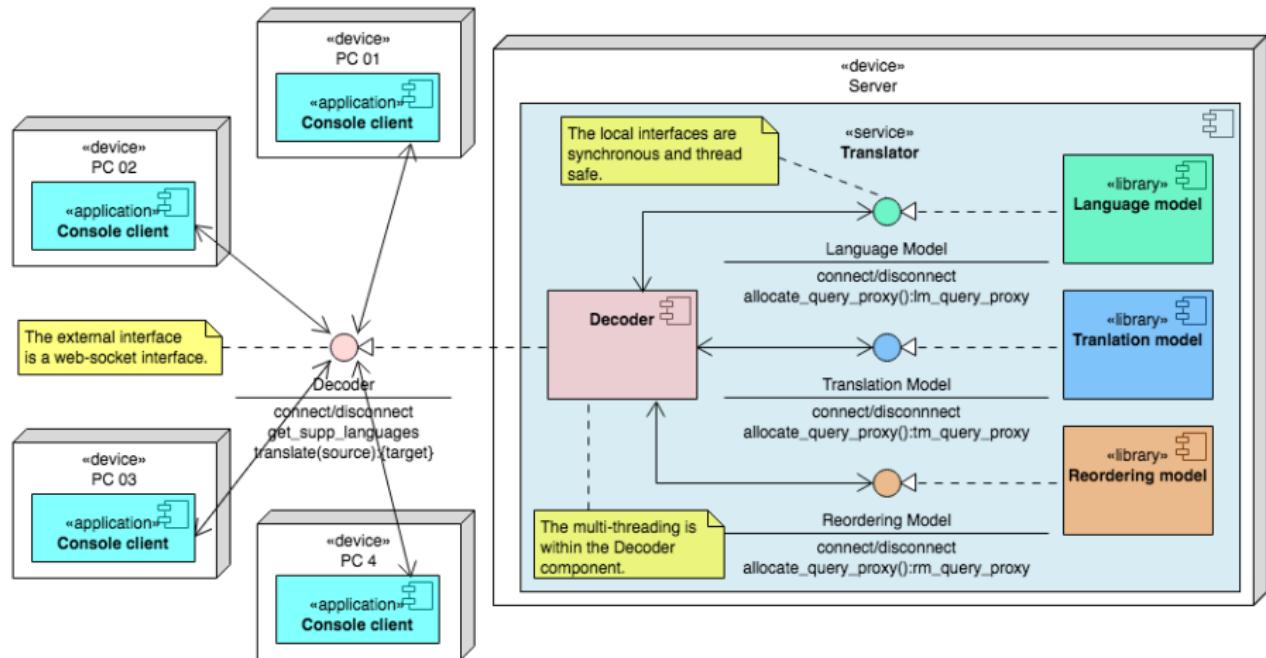
- UvA • ivan.zapreev@uva.nl
- UvA • c.monz@uva.nl

Retrospective

Statistical Machine Translation



Retrospective: Previous 2 deliveries



Functionality, Performance, Profiling, Documentation

Present days

Month 18 delivery

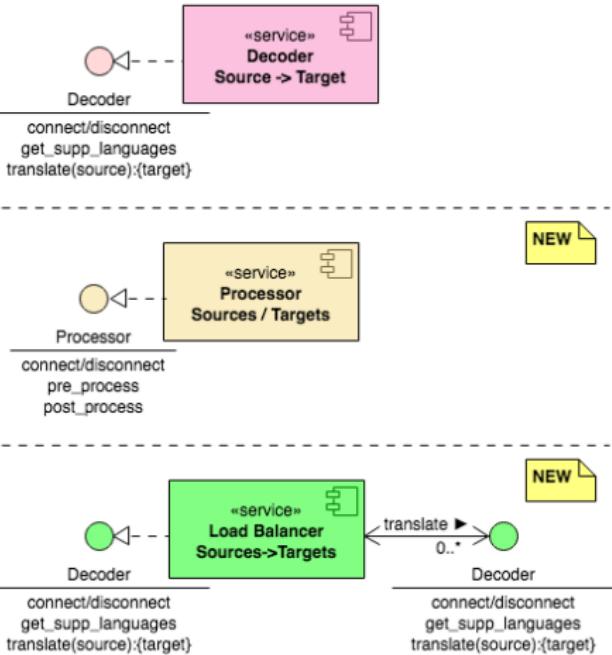
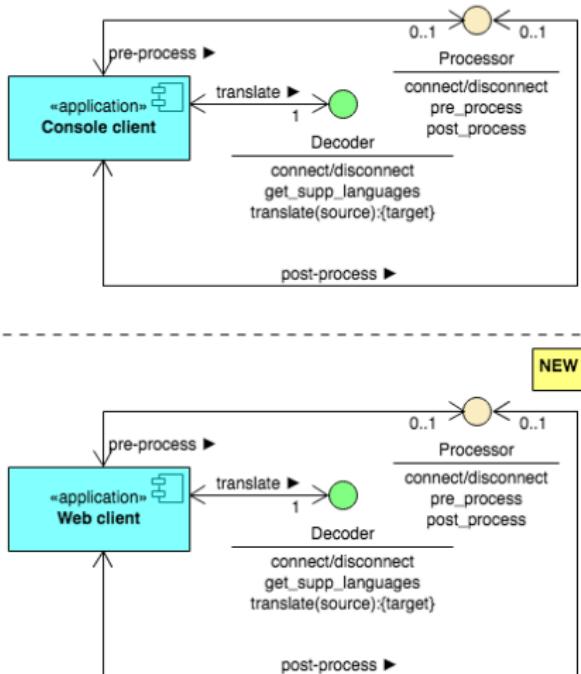
Expected:

- S1, WP1:
 1. Web front end
 2. Load balancing
 3. Search lattice
 4. Complete infr.
- R3:
 - Report

Additional:

- Job priorities
- Deployments
- Extended docs
- Design diagrams
- Validation, testing
- Def. pre/post scripts

Delivered infrastructure components



An example deployment

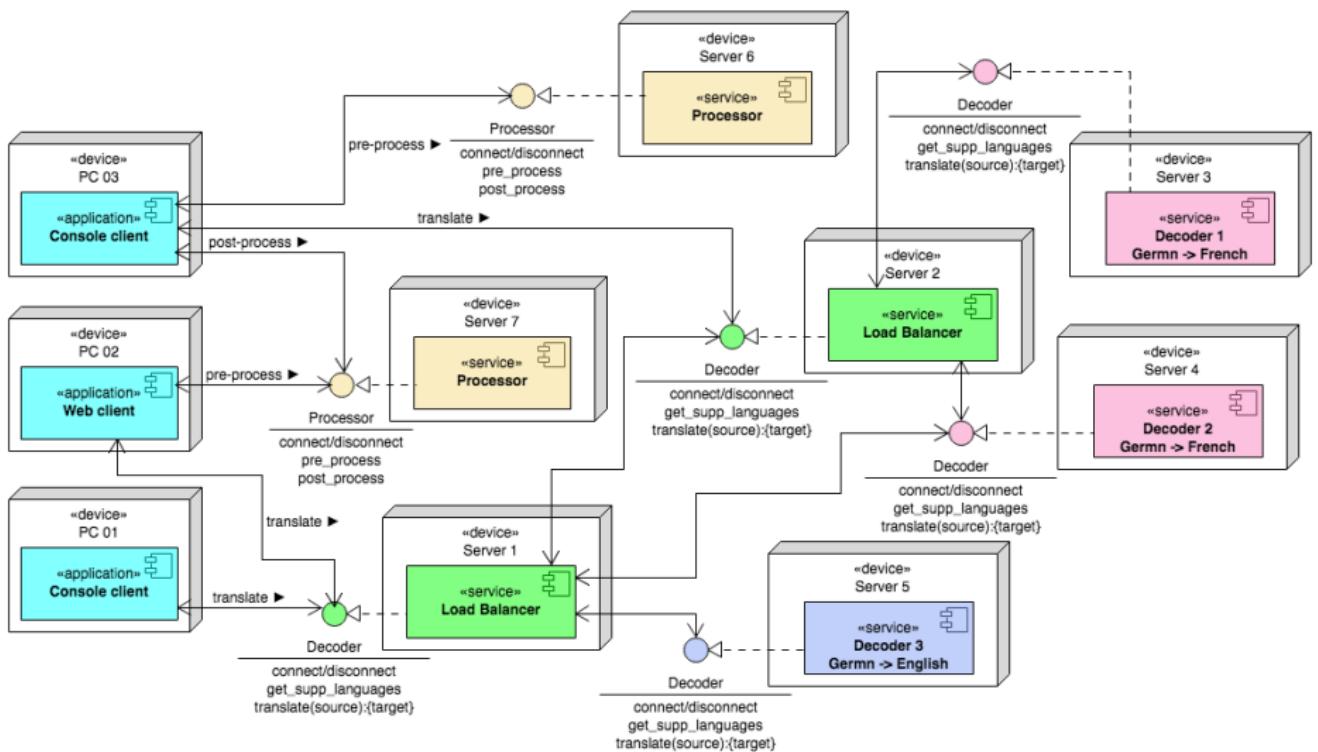


Table of Contents

- Console Client
- Web Client
- Translation Server
- Load Balancer
- Text Processor
- Possible Deployments
- Demo Infrastructure
- Testing and Robustness
- Conclusions

Console client

```
Drs-MacBook-Pro:build zapreevis$ bpbd-client
USAGE: -----
USAGE: |           The translation client application      :)\_ /(: |
USAGE: |           Software version 1.4                  {(@)\v(@) } |
USAGE: |           The Owl release.                   {|\~-\_~|} |
USAGE: |           Copyright (C) Dr. Ivan S Zapreev, 2015-2016 {/^'`'^\} |
USAGE: | _____m-m_____
USAGE: |           This software is distributed under GPL 2.0 license |
USAGE: |           (GPL stands for GNU General Public License) |
USAGE: |           The product comes with ABSOLUTELY NO WARRANTY. |
USAGE: |           This is a free software, you are welcome to redistribute it. |
USAGE: |           Running in 64 bit mode! |
USAGE: |           Build on: Oct 17 2016 19:50:52 |
USAGE: -----
PARSE ERROR:
                    Required arguments missing: output-file, input-lang, input-file

Brief USAGE:
  bpbd-client [-d <error|warn|usage|result|info|info1|info2|info3>] [-c]
              [-s <the translation priority>] [-l <min #sentences per
              request>] [-u <max #sentences per request>] [-p
              <post-processor uri>] [-t <server uri>] [-r <pre-processor
              uri>] [-o <target language>] -O <target file name> -i
              <source language> -I <source file name> [--] [--version]
              [-h]

For complete USAGE and HELP type:
  bpbd-client --help
```

New: Internal design, Pre/Post-processor, Job priorities

Job priorities

Assign each translation job a priority value.

- Priority
 - Integer value
 - Natural order
 - Default is 0
- Queue
 - C++ standard class
 - First In First Out
- Infrastructure-wide support

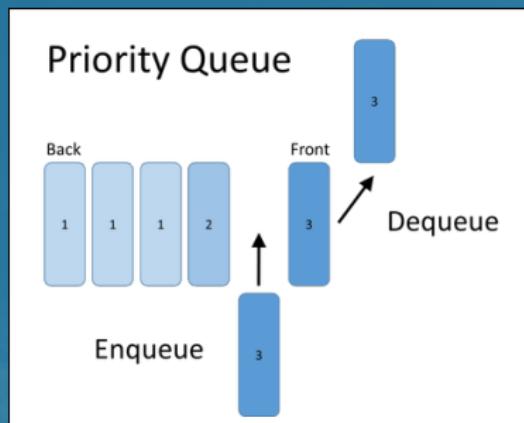
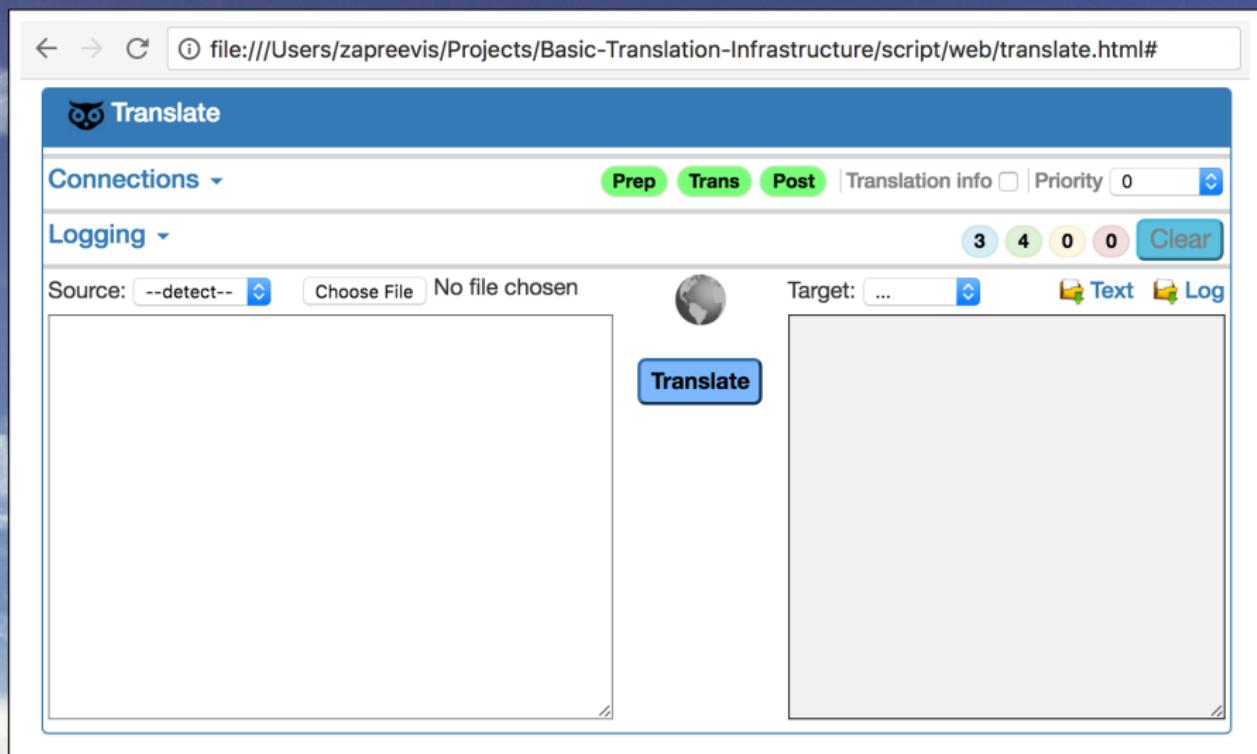


Table of Contents

- Console Client
- **Web Client**
- Translation Server
- Load Balancer
- Text Processor
- Possible Deployments
- Demo Infrastructure
- Testing and Robustness
- Conclusions

Web client: Technologies



WebSockets, JSON, JavaScript, jQuery 2.2.4, Bootstrap 3.3.6

Web client: Small demo



Table of Contents

- Console Client
- Web Client
- Translation Server
- Load Balancer
- Text Processor
- Possible Deployments
- Demo Infrastructure
- Testing and Robustness
- Conclusions

Translation server

```
Drs-MacBook-Pro:build zapreevis$ bpbd-server
USAGE: -----
USAGE: |           The translation server application      :)\_\_/(:
USAGE: |           Software version 1.4                  {(@)v(@)}
USAGE: |           The Owl release.                   {|~~ -~|}
USAGE: |           Copyright (C) Dr. Ivan S Zapreev, 2015-2016{/^\^\^}\_
USAGE: | _____m-m_____
USAGE: |           This software is distributed under GPL 2.0 license
USAGE: |           (GPL stands for GNU General Public License)
USAGE: |           The product comes with ABSOLUTELY NO WARRANTY.
USAGE: |           This is a free software, you are welcome to redistribute it.
USAGE: |           Running in 64 bit mode!
USAGE: |           Build on: Oct 31 2016 10:15:57
USAGE: | _____m-m_____
PARSE ERROR:
    Required argument missing: config

Brief USAGE:
  bpbd-server [-d <error|warn|usage|result|info|info1|info2|info3>] -c
               <server configuration file> [--] [--version] [-h]

For complete USAGE and HELP type:
  bpbd-server --help
```

New: Internal design, Job priorities, Search Lattice Generation

Translation server: Search Lattice

Purpose:

Search Lattice → Discriminative Training → Reduce Error

Discriminative Training:

Tune lambda weights of model features

Reduce Error:

Increase the BLEU scores - translation quality

How-To:

Compile in tuning mode; Configure parameters; Use lattice combination script

Table of Contents

- Console Client
- Web Client
- Translation Server
- Load Balancer
- Text Processor
- Possible Deployments
- Demo Infrastructure
- Testing and Robustness
- Conclusions

Load Balancer

Purpose:

- Distribute job load
- Aggregate decoders

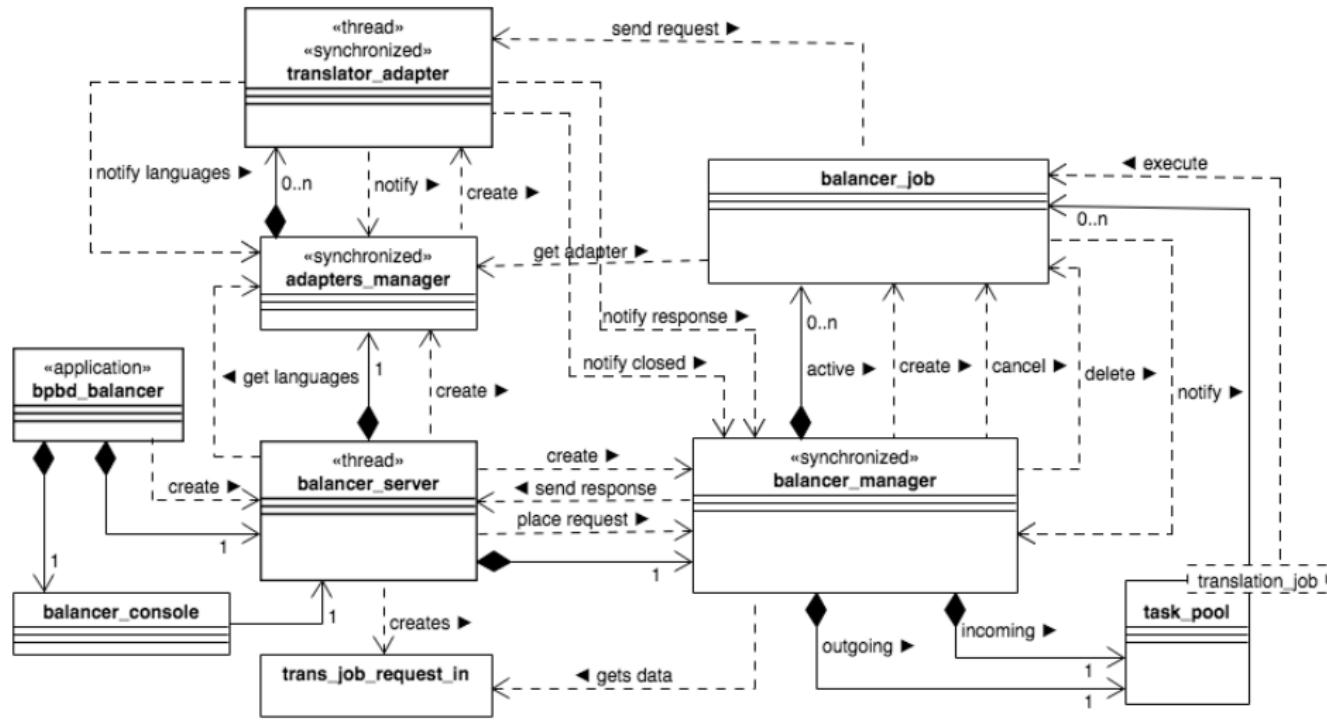
Features:

- Multi-threaded
- Job priorities
- Auto-reconnect
- Load weights



Load weight is a positive integer indicating the work capacity of the translation server relative to the other ones.

Load Balancer: Internal Design



Balancer console; Auto re-connect; Duplex communications

Table of Contents

- Console Client
- Web Client
- Translation Server
- Load Balancer
- **Text Processor**
- Possible Deployments
- Demo Infrastructure
- Testing and Robustness
- Conclusions

Text Processor

Purpose: Transformation

- **Pre:** Human → Decoder
- **Post:** Human ← Decoder

Features:

- Multi-threaded
- Job priorities
- Integration with 3rd party tools



Pre: De-capitalize, tokenize, split sentences, ...

Post: Truecase, de-tokenize, restore structure, ...

Pre-processor script

```
#Pre-processor script call template
#The <WORK_DIR> the work directory where the files are stored
#The <JOB_UID> the unique processor job identifier.
#The <LANGUAGE> the source language name

pre_call_tmpl=./pre_process.sh <WORK_DIR> <JOB_UID> <LANGUAGE>
```

- Can be anything executable
- Fixed program interface
- Default shell script:
 - Copies inputs to outputs
 - Detects inputs as German
 - Has extended help

Post-processor script

```
#Post-processor script call template
#The <WORK_DIR> the work directory where the files are stored
#The <JOB_UID> the unique processor job identifier.
#The <LANGUAGE> the target language name

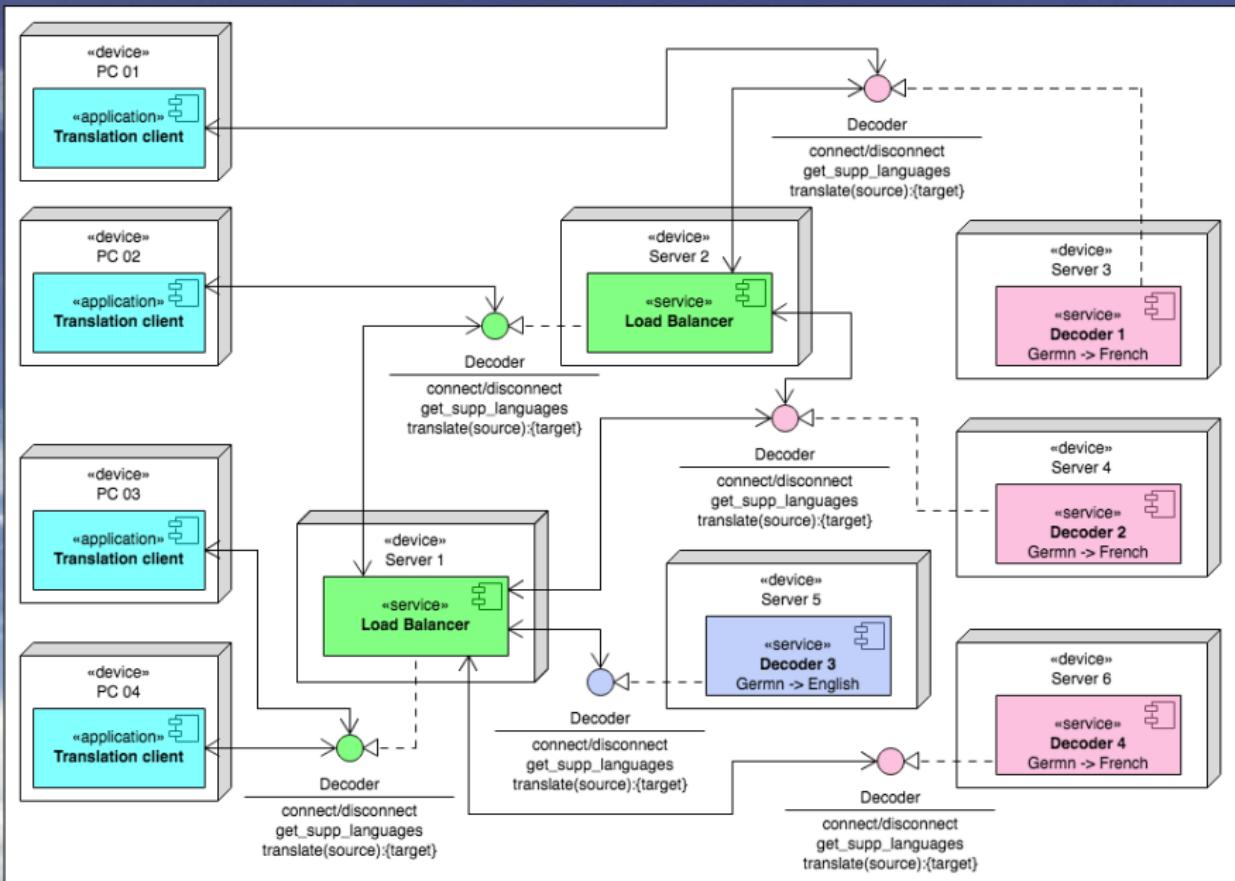
post_call_templ=./post_process.sh <WORK_DIR> <JOB_UID> <LANGUAGE>
```

- Can be anything executable
- Fixed program interface
- Default shell script:
 - Copies inputs to outputs
 - Has extended help

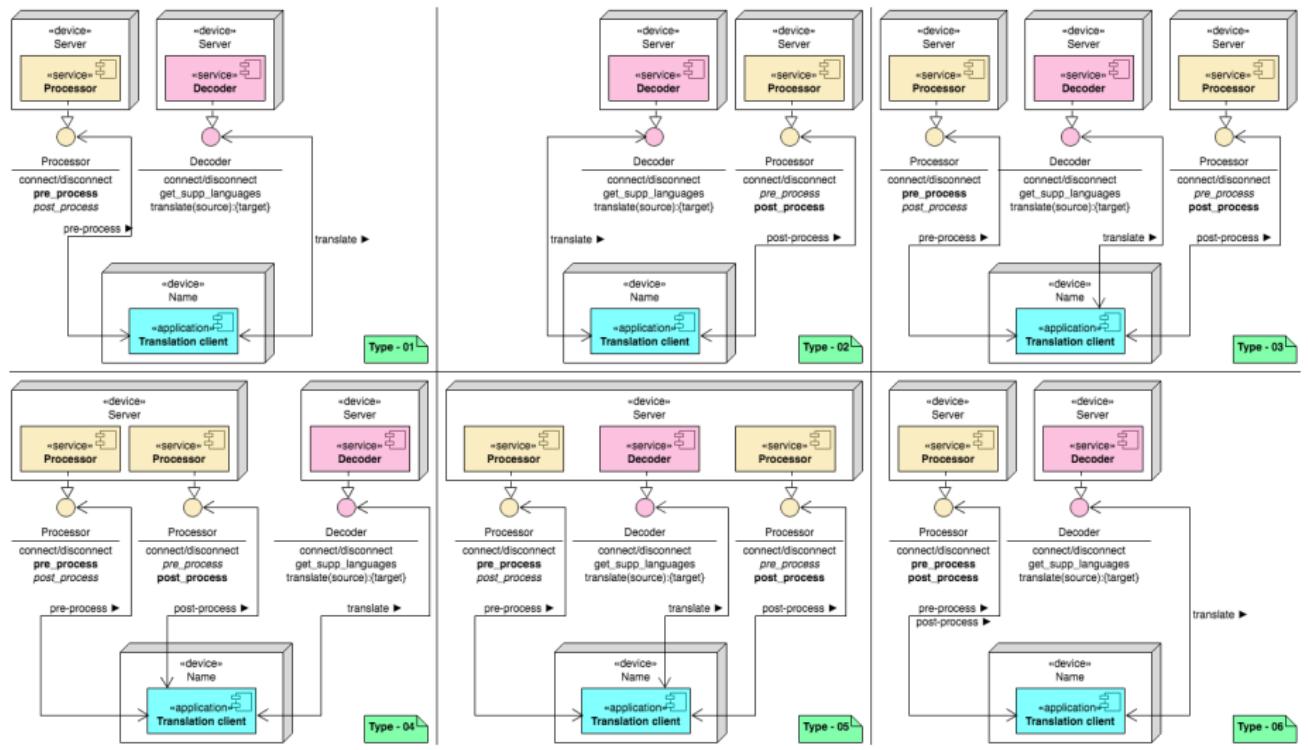
Table of Contents

- Console Client
- Web Client
- Translation Server
- Load Balancer
- Text Processor
- **Possible Deployments**
- Testing and Robustness
- Conclusions

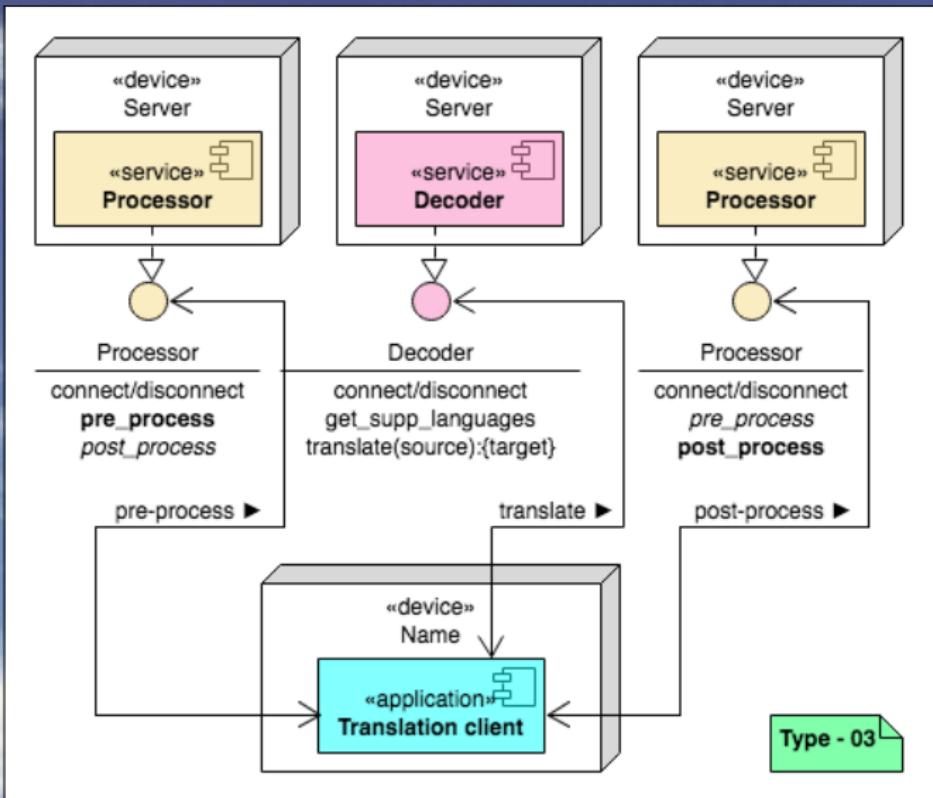
Possible deployments: Balancer



Possible deployments: Processor



Possible deployment: Type 03



Possible deployment: Type 06

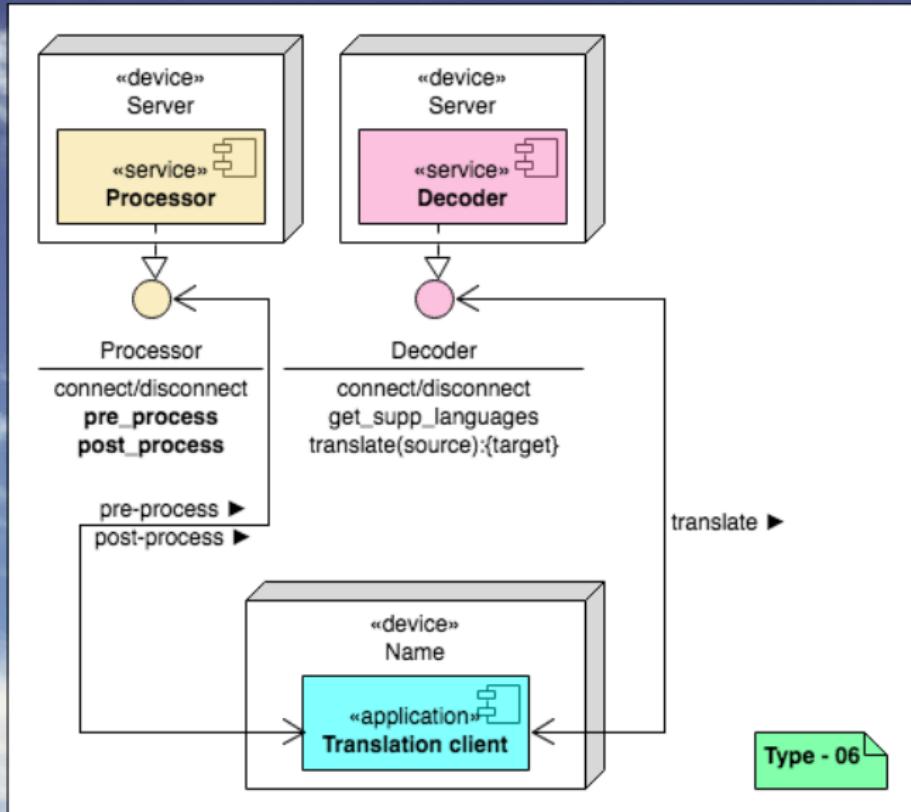
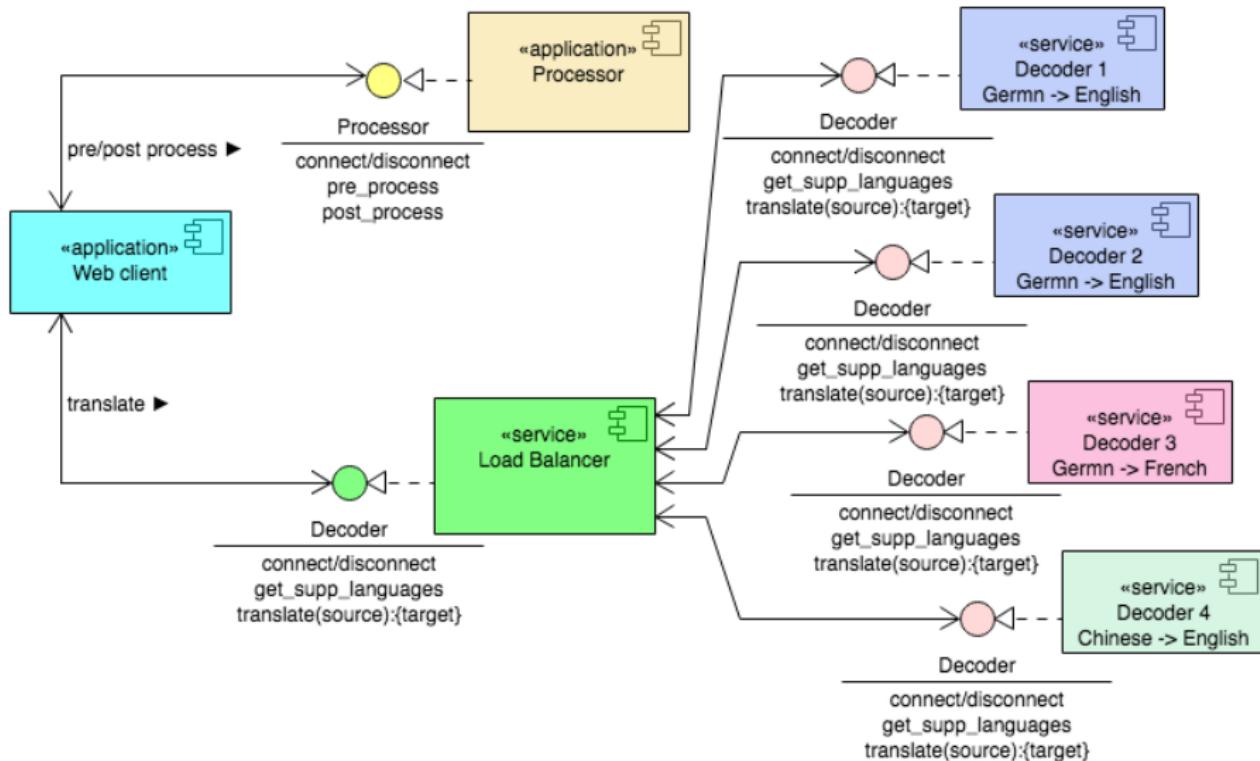


Table of Contents

- Console Client
- Web Client
- Translation Server
- Load Balancer
- Text Processor
- Possible Deployments
- **Demo Infrastructure**
- Testing and Robustness
- Conclusions

Demo Infrastructure: Deployment



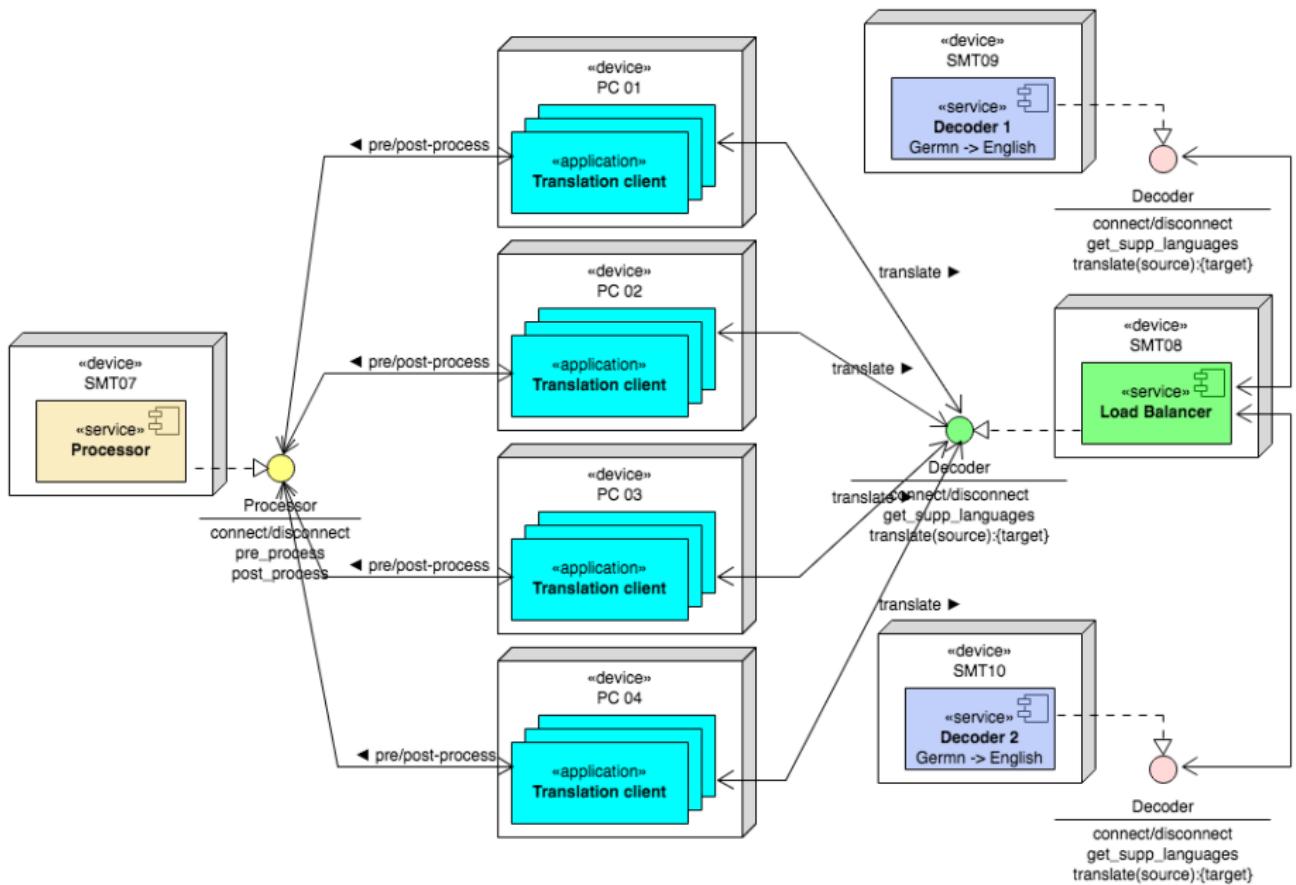
Demo Infrastructure: Demo



Table of Contents

- Console Client
- Web Client
- Translation Server
- Load Balancer
- Text Processor
- Possible Deployments
- Demo Infrastructure
- **Testing and Robustness**
- Conclusions

Stress testing infrastructure



Testing and Robustness

- Stress testing

- 4 server machines
 - SMT07: Processor - 20 threads
 - SMT08: Balancer - 10 threads
 - SMT09: Decoder 1 - 40 threads
 - SMT10: Decoder 2 - 40 threads
- 4 client machines
 - 400 clients per machine
 - 3.000 sentences per client
- #connections: $4 \times 400 = 1.600$
- #tasks: $4 \times 400 \times 3.000 = 4.800.000$
- Test time $\approx 36 - 48$ hours

- Valgrind profiling

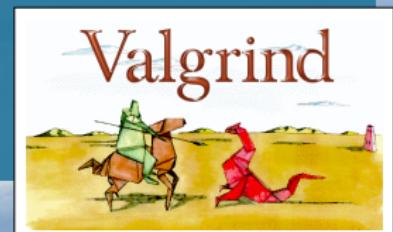
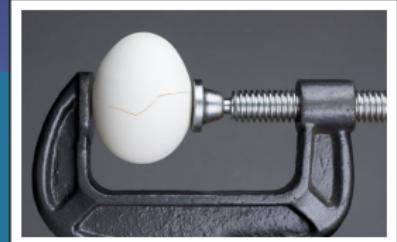


Table of Contents

- Console Client
- Web Client
- Translation Server
- Load Balancer
- Text Processor
- Possible Deployments
- Demo Infrastructure
- Testing and Robustness
- **Conclusions**

Conclusions

A complete, on time, and robust delivery.

As expected:

- Web front end
- Load balancer
- Text processor
- Search lattice
- Complete infr.
- Project report

Additional:

- Job priorities
- Deployments
- Extended docs
- Design diagrams
- Validation, testing
- Def. pre/post scripts

Future plans

- Complete tuning framework
- Better BLEU scores, e.g.
 - Chinese → English
 - 23 → 36
- Neural LM integration
- Pre/Post-processing examples
- Compare Performance with Oister.

Discussion



Thank you all!

Wednesday
July 1