Lessons Learnt from Developing & Maintaining the World's Largest* CP Model Using MiniZinc

Presented by Erik Cervin-Edin

Content

- Quick background on Ericsson and RAN networks
- Using combinatorial optimization in product configuration
- Developing, executing & maintaining very large CP models

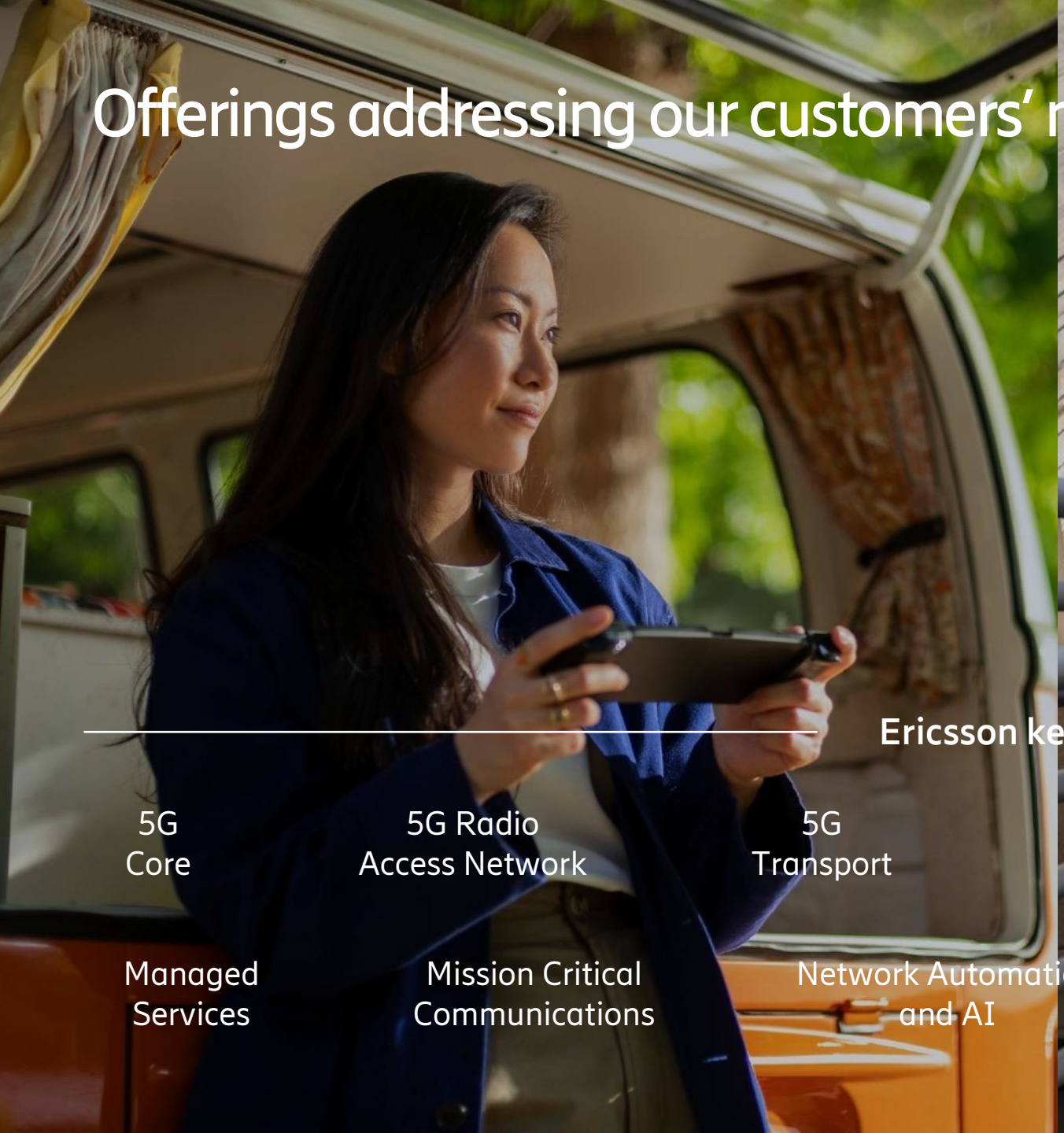
Erik @ Ericsson

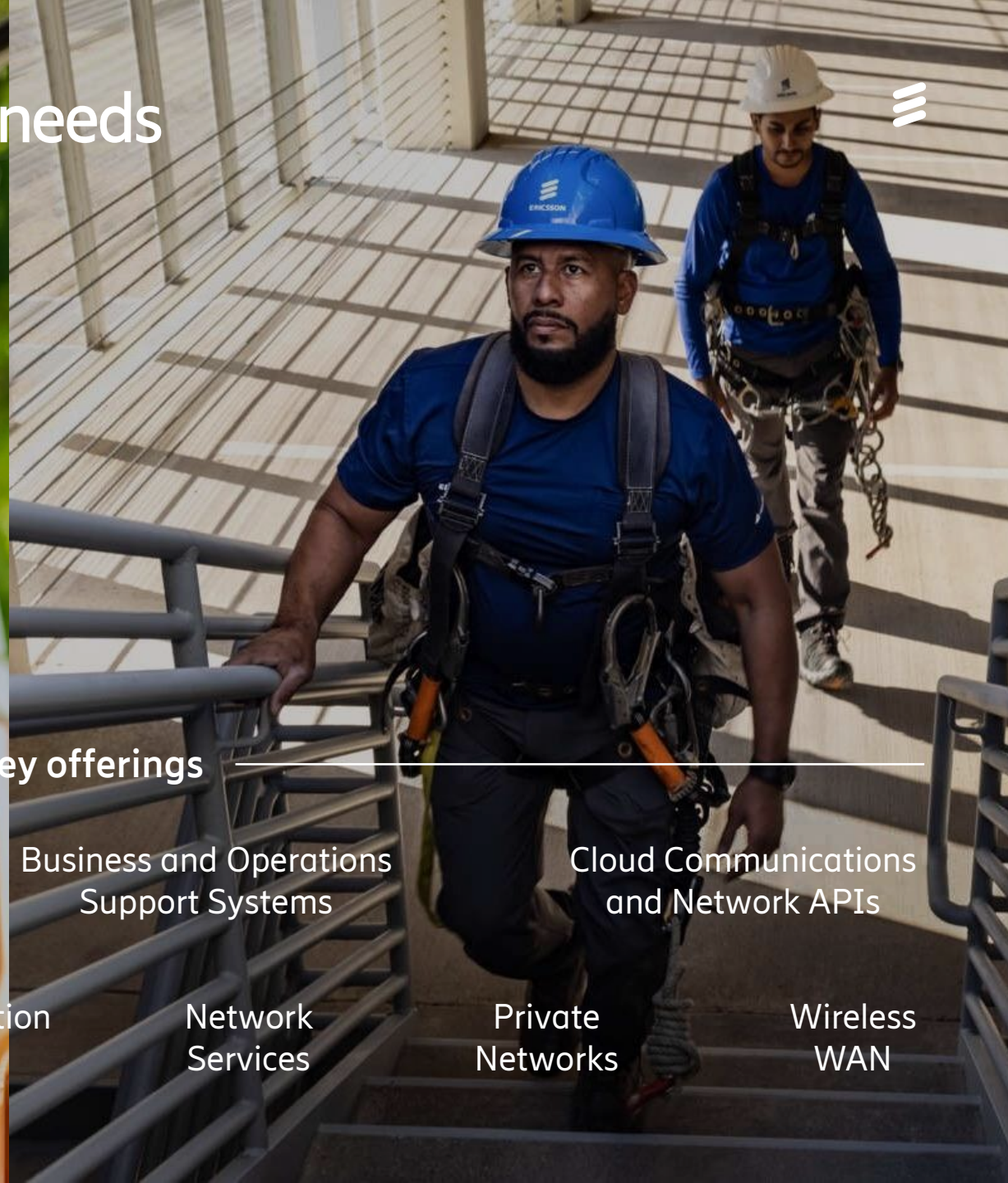
- Erik Cervin Edin
- Software Developer @ Ericsson
- Computer Science, Uppsala University
- Ericsson since Feb 2023

https://github.com/CervEdin

https://linkedin.com/in/erikcervinedin

Offerings addressing our customers' needs

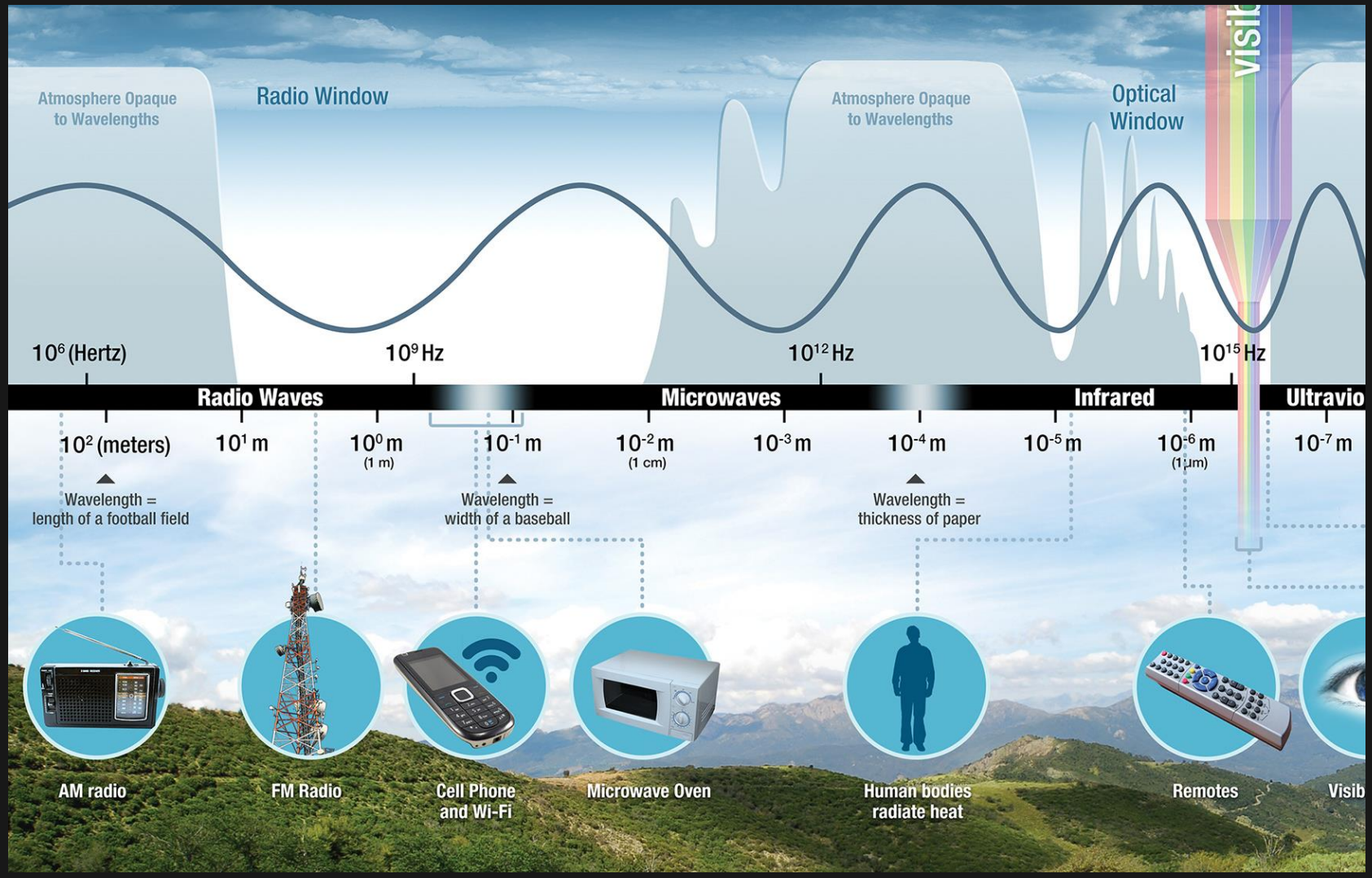




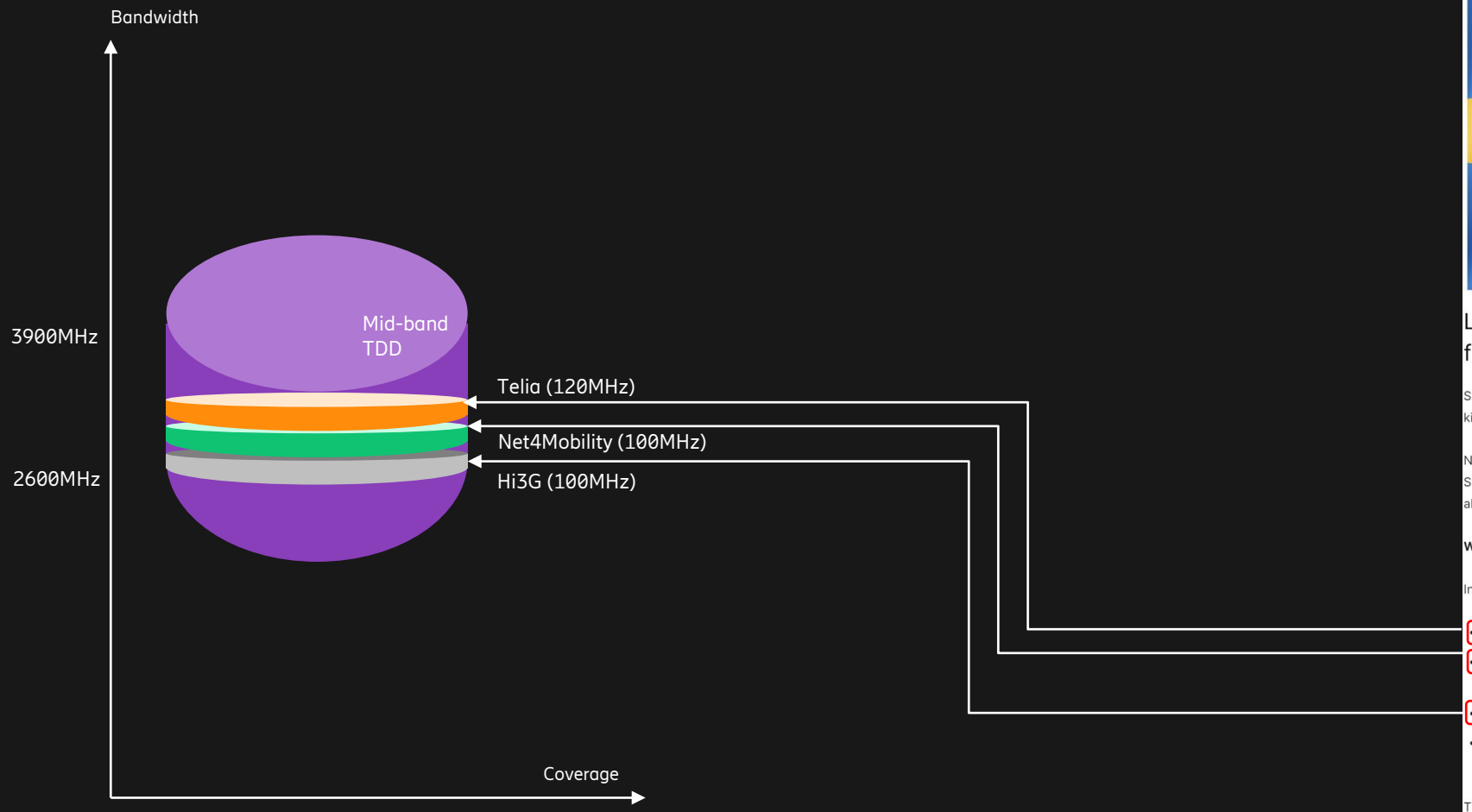
Ericsson key offerings

- 5G Core
- 5G Radio Access Network
- 5G Transport
- Business and Operations Support Systems
- Cloud Communications and Network APIs
- Managed Services
- Mission Critical Communications
- Network Automation and AI
- Network Services
- Private Networks
- Wireless WAN

Electromagnetic spectrum



Spectrum allocation



5G & Beyond | RAN | Access | News

Sweden completes spectrum auction in one day

By **Annie Turner** - 22 September 2023

Share

Licences were up for grabs in the 900MHz, 2.1GHz and 2.6GHz frequency ranges

Sweden's Post and Telecom Authority (PTS) announced the conclusion of its latest spectrum auction which kicked off on Wednesday.

Nordic telecom companies Tele2, Telenor Sweden and Telia Co have all acquired licences in the latest Swedish spectrum auction. They collectively invested SEK3.03 billion (€254.68 million) for spectrum allocations in the 900MHz, 2.1GHz and 2.6GHz auction.

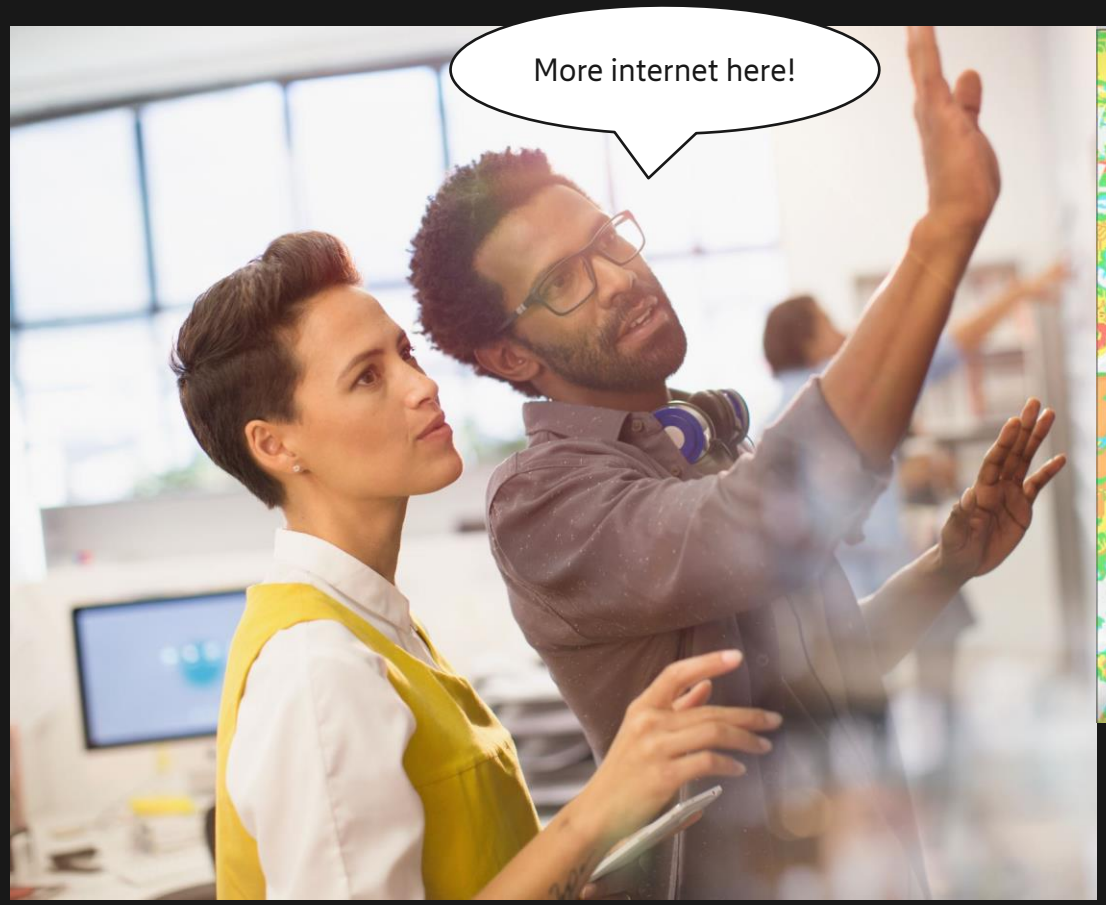
Who got what

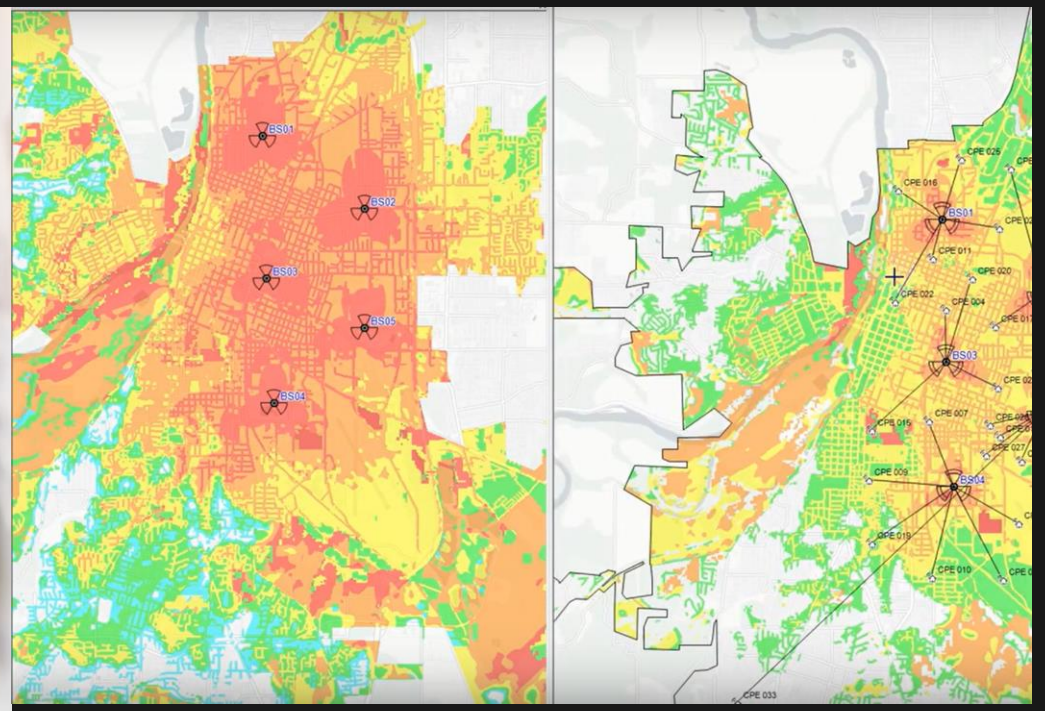
In a statement, PTS noted all 320 MHz at 3.5 GHz was assigned. Full allocations are as follows:

- Telia secured 120MHz (3500-3620 MHz) for SEK760.25 million SEK (€75 million)
- Net4Mobility (the joint venture between Tele2 and Telenor Sweden) won 100MHz (3620-3720MHz) for SEK665.5 million (€65 million)
- Hi3G secured 100MHz (3400-3500 MHz) at SEK491.25 million (€48 million)
- Teracom Group, which took over Net1 in 2019, won all the 80MHz on offer in the 2.3 GHz band for a total of SEK400 million (€40 million)

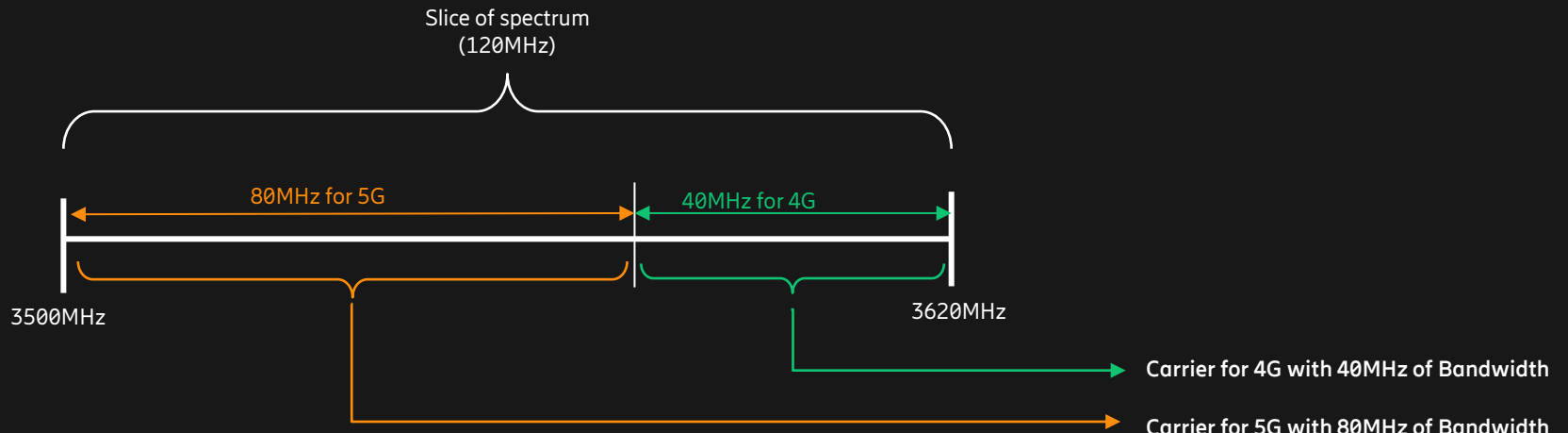
The four 3.5 GHz licences will be valid for a period of 25 years, from 20 January, 2021 to 31 December, 2045.

The network plan





Output of Planning Activity

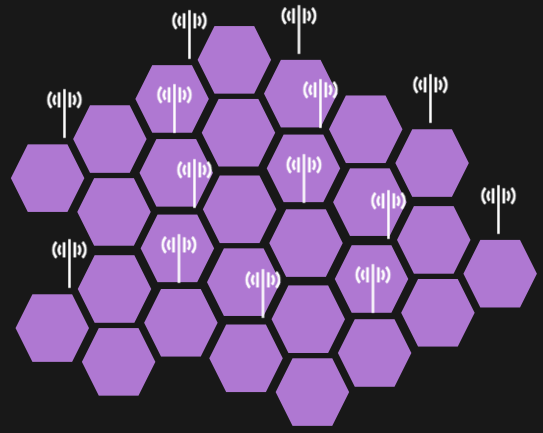


Option-1
- Full 120MHz for LTE (4G)

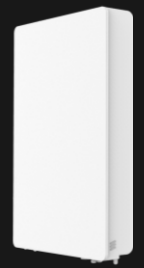
Option-2
- Full 120MHz for NR (5G)

Option-3
- Mix of 4G & 5G
  - 60MHz for each
  - 40MHz for LTE & 80MHz for NR
  - And so on.....

Number of Sites

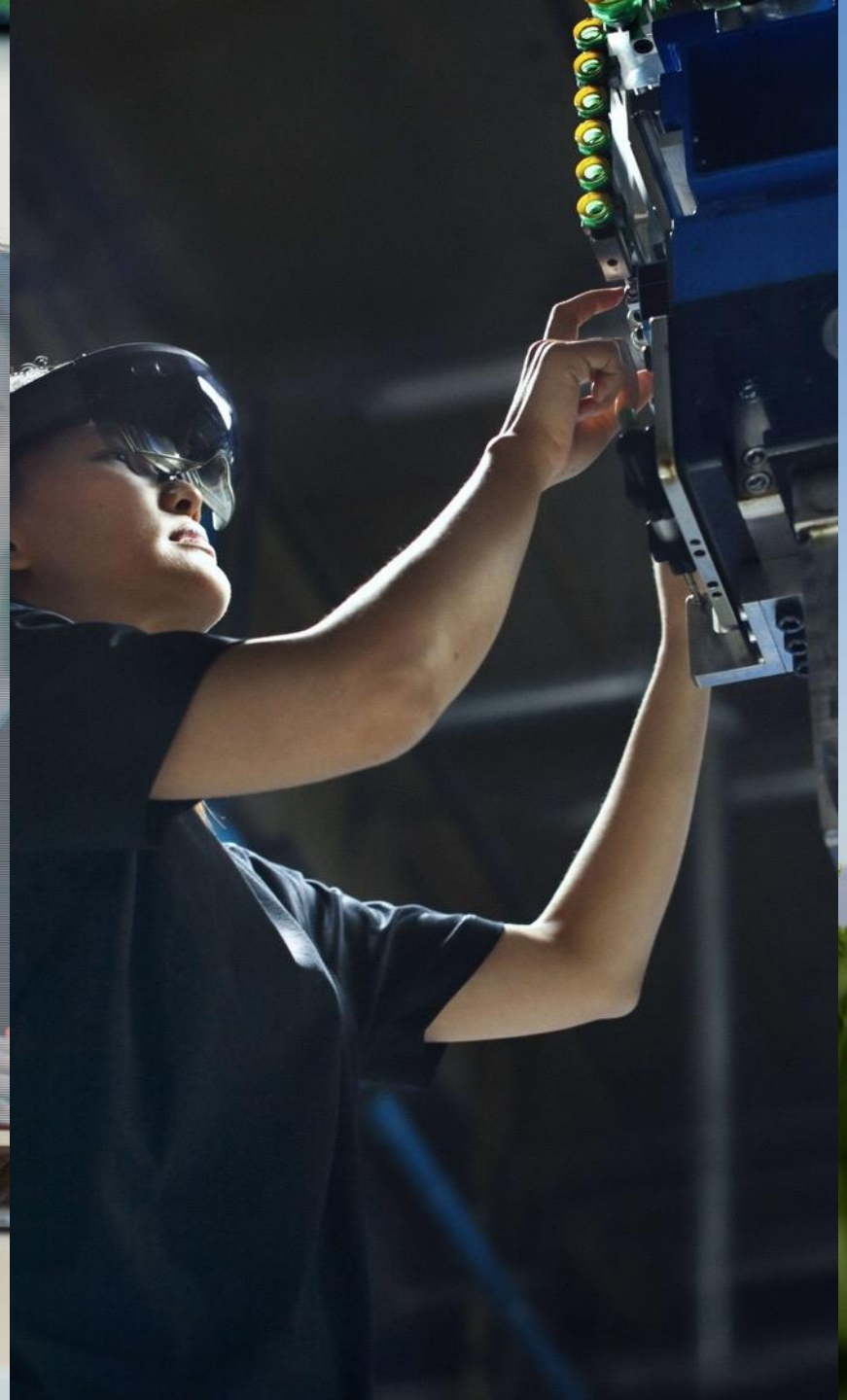


Radio ?? (3500-3700MHz)



Radio ?? (3410-3800MHz)

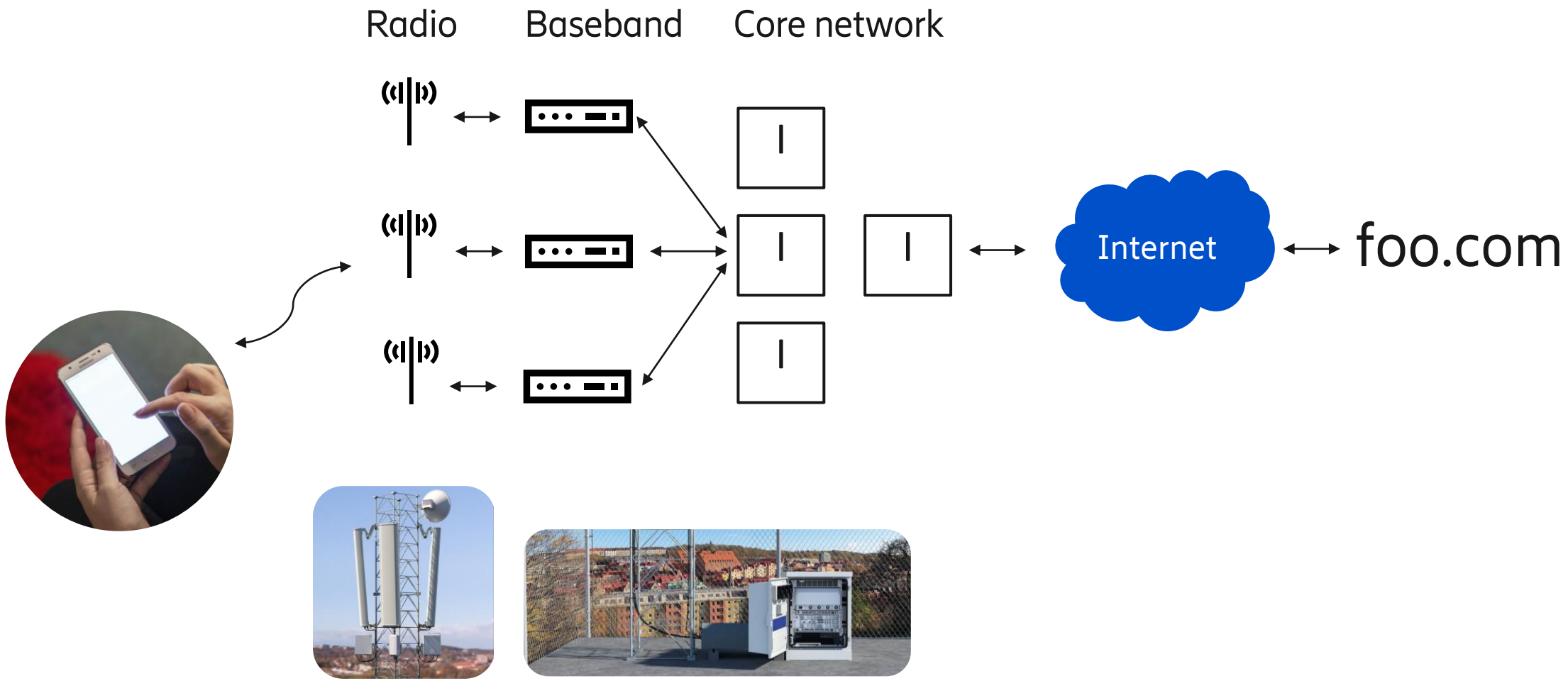
Radio ?? (3450-3800MHz)



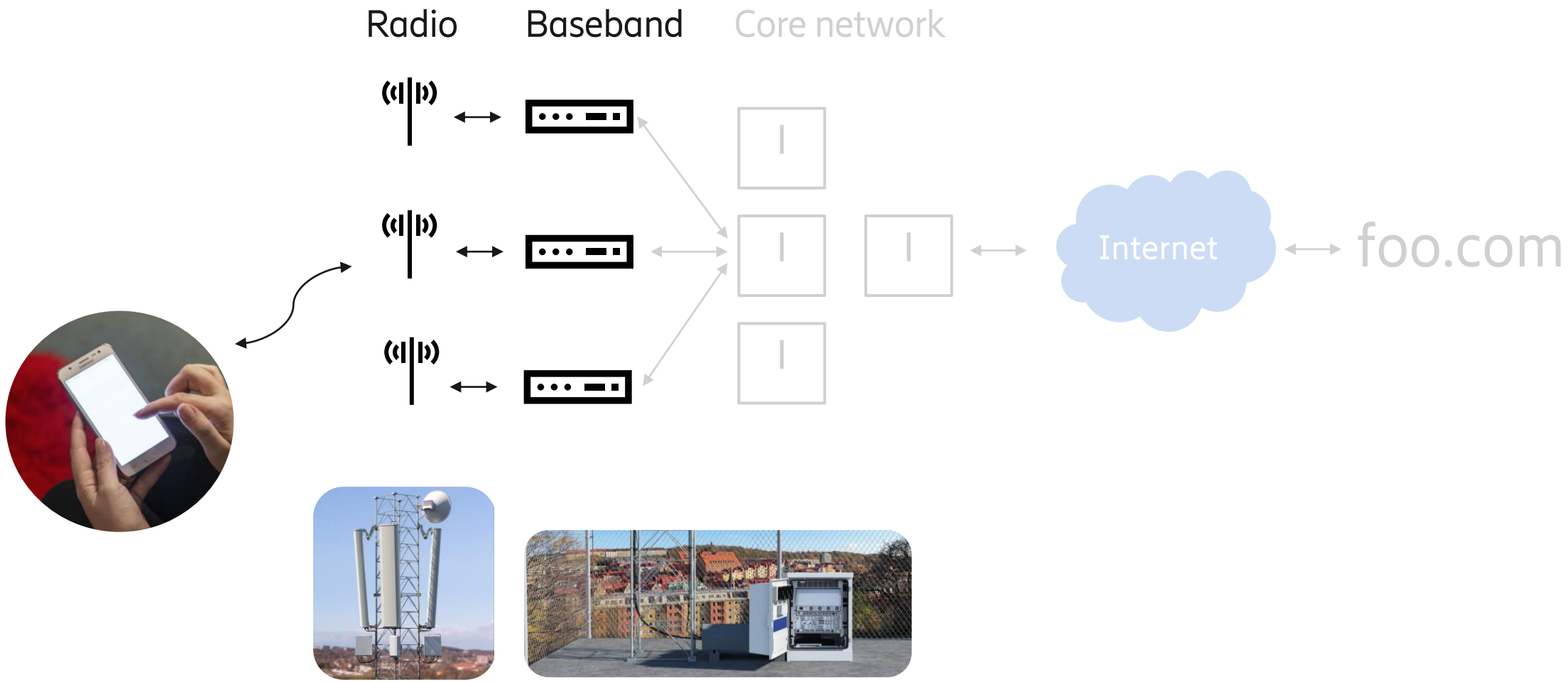
Product Configuration @ Ericsson

- Aid sales & support
- Configuration engines customize products to meet needs – like buying a couch

Radio Access Network (RAN) Overview

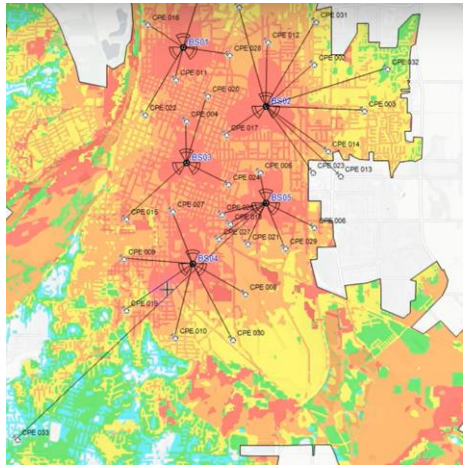


Radio Access Network (RAN) Overview



Product configuration – The requirements

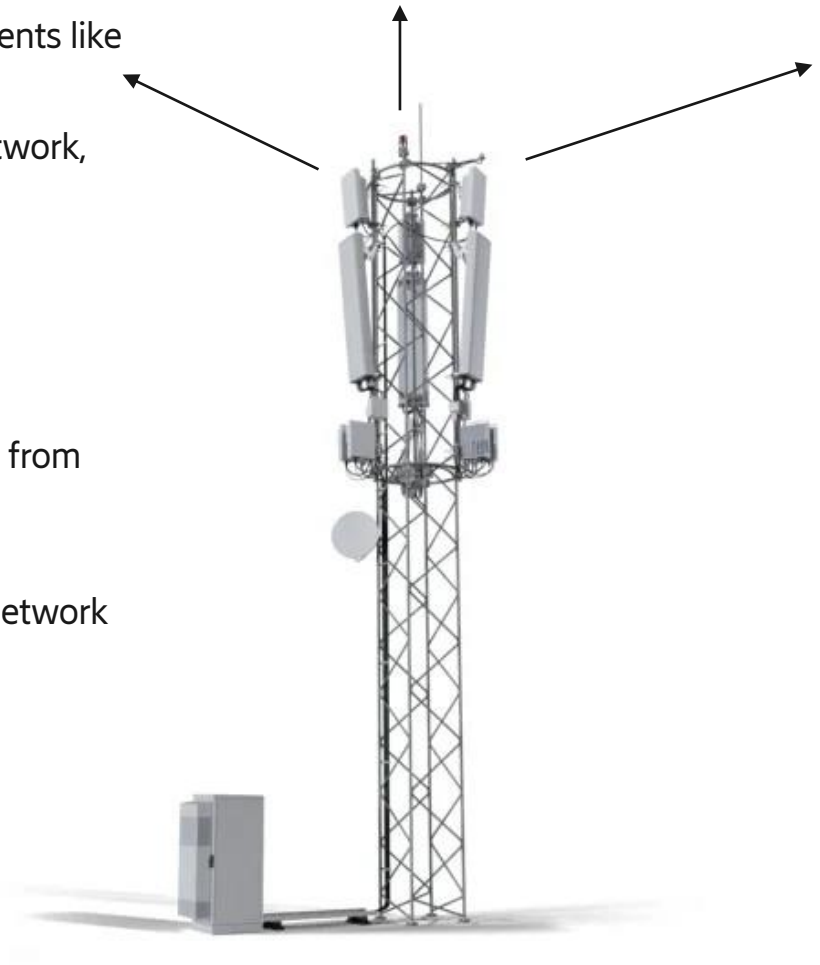
- Site 1
  - GSM carrier
  - 3G carrier
  - 2 x 5G carrier
- Site 2
  - 2 x 4G carrier
  - 5G carrier
- Site 3
  - GSM carrier
  - LTE carrier
- …



Product configuration – The site

1+ radio solutions

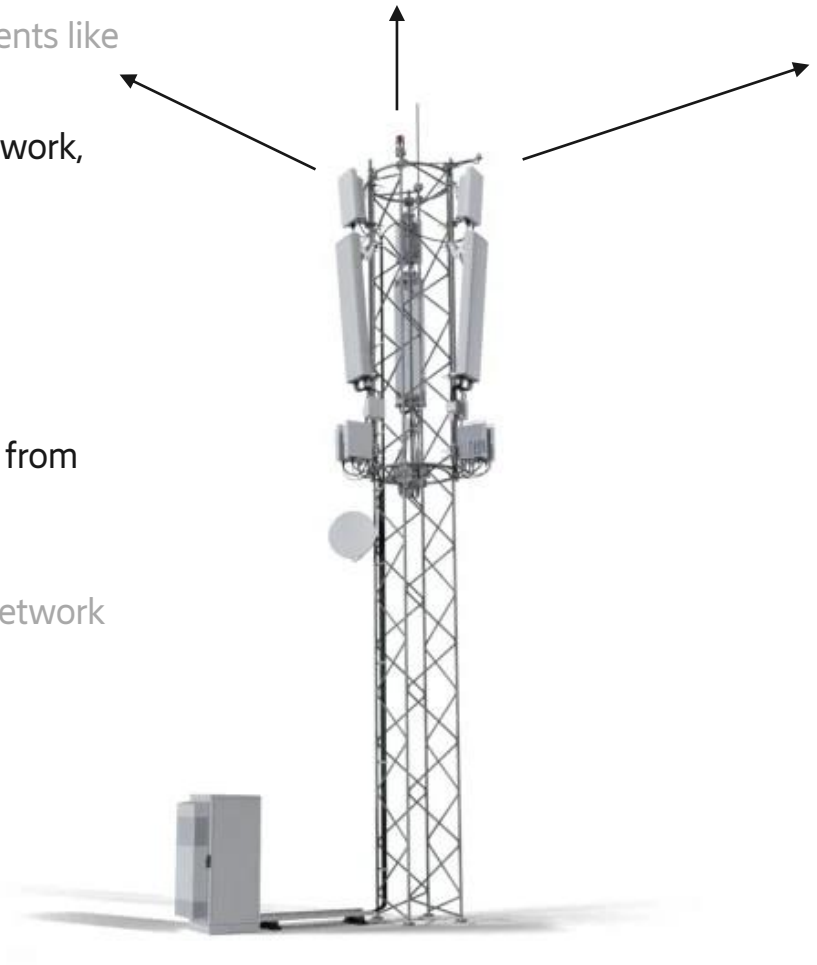
- **Antenna System**: responsible for transmitting and receiving radio signals. It includes components like antennas, cables, and connectors.
- **Carrier:** range of frequencies allocated for transmitting and receiving signals on a wireless network, typically defined by its center frequency and bandwidth. (ephemeral)
- **RF Port:** interface that connects radio to antennas, split RX⬇/TX⬆ traffic.
- **Radio:** wireless communication component that transmits and receives radio signals.
- **CPRI Ports:** interface that connects radio to baseband, ⬇⬆ traffic.
- **Baseband:** network that handles the lower frequency signals, after they have been converted from radio frequencies (RF) by an antenna and receiver. It performs tasks such as switching, traffic management, timing, baseband processing, and radio interfacing.
- **RAN Compute**: This refers to the computing resources required to support the Radio Access Network (RAN), such as baseband units or virtualized RAN functions.
- **Enclosure**: physical housing that protects the radio equipment
- **Power Supply**: provides the electrical power to operate the radio solution
- …



Product configuration – The site

1+ radio solutions

- **Antenna System**: responsible for transmitting and receiving radio signals. It includes components like antennas, cables, and connectors.
- **Carrier:** range of frequencies allocated for transmitting and receiving signals on a wireless network, typically defined by its center frequency and bandwidth. (ephemeral)
- **RF Port:** interface that connects radio to antennas, split RX⬇/TX⬆ traffic.
- **Radio:** wireless communication component that transmits and receives radio signals.
- **CPRI Ports:** interface that connects radio to baseband, ⬇⬆ traffic.
- **Baseband:** network that handles the lower frequency signals, after they have been converted from radio frequencies (RF) by an antenna and receiver. It performs tasks such as switching, traffic management, timing, baseband processing, and radio interfacing.
- **RAN Compute**: This refers to the computing resources required to support the Radio Access Network (RAN), such as baseband units or virtualized RAN functions.
- **Enclosure**: physical housing that protects the radio equipment
- **Power Supply**: provides the electrical power to operate the radio solution
- ...



Product configuration – Which is best?

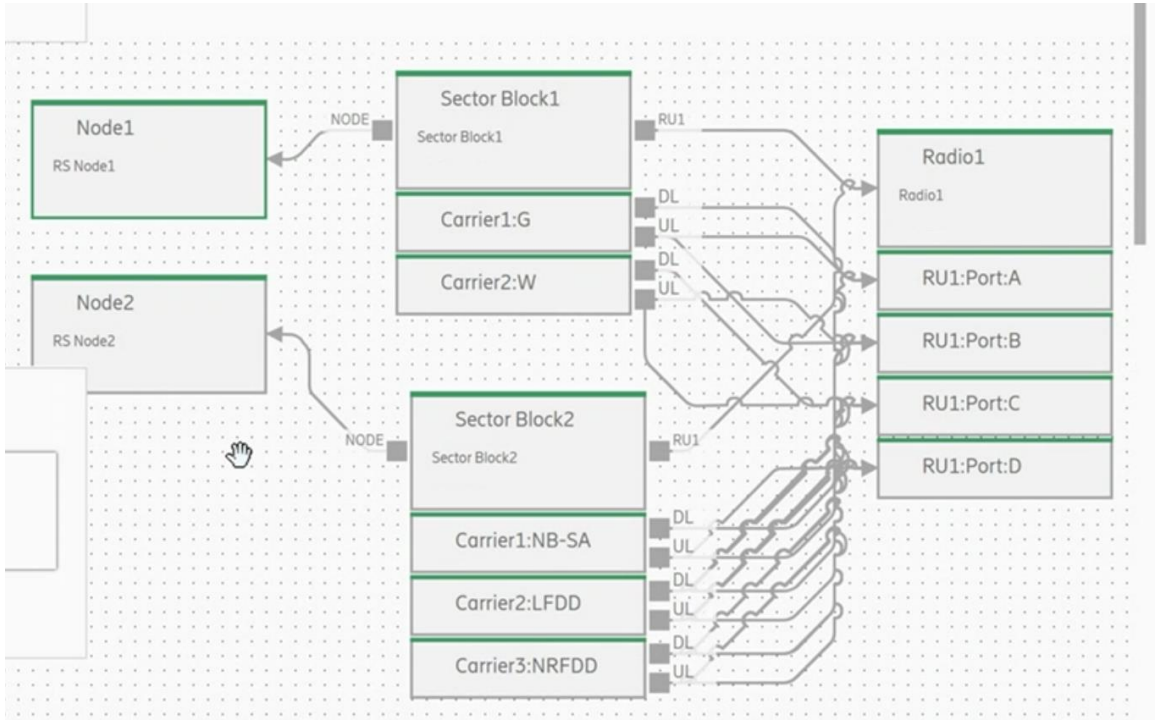
Lexicographic Optimization

- Minimize radio equipment
  - number of radios
  - …
- Optimize
  - Output power
  - Weight
  - Size
  - Other customer desires?
- The objective function is subjective!

Product configuration – Radio solution

Resource allocation - a combinatorial optimization problem

- Radio Solution:
  - Carriers ⇆ Radios ↔ Basebands
- Challenge:
  - CSP: Allocating components
  - COP: Minimize waste
- Solution:
  - Bin-packing (e.g. connecting cables)
  - ~50 table constraints
    - channeling/side-constraints (e.g. HW specific capabilities)
  - Linear constraints (capacities)



Given a set C of carriers, select a set R of radio units and map every carrier $c \in C$ to a radio $r \in R$ such that r meets all demands of c mapped to r and the capacities of r are not exceeded.

Structuring a large CP model

Modularizing a MiniZinc model

The core

```
 1
 2 % Declarations (parameters, functions etc.)
 3 include "model/types.mzn";
 4 include "model/enums.mzn";
 5 include "model/data_tables.mzn";
 6 include "model/problem.mzn";
 7 include "model/utils.mzn";
 8
 9 % Decision variables
10 include "model/decision_vars.mzn";
11
12 % Constraints
13 include "model/constraints_core.mzn";
14 include "model/constraints_special.mzn";
15
16 % Improving constraints
17 include "model/constraints_improving.mzn";
18
19 % Search annotations
20 include "model/search.mzn";
21
22 % Input sanity checks
23 include "model/verify_input.mzn";
```

- Declarations
  - Variables
    - Constraints
      - Improving constraints
      - Search annotations
      - Sanity checks (assertions)
- Easier to debug!
- Easier to maintain!

Modular configurations

The default configuration

```
{
  "solver": "or-tools",
  "free-search": true,
  "model": [
    "model.mzn",
    "model/input/default.mzn",
    "model/objective/default.mzn",
    "model/output/json.mzn",
    "model/solve/minimize.mzn",
    "aux_tables.mzn",
    "sets.mzn"
  ],
  "data": [
    "data_ept.dzn",
    "data_static.dzn",
    "data_tables_ept.dzn",
    "enums.dzn",
    "enums_static.dzn"
  ]
}
```

- All files included in model.mzn
- Modularized input
  - multiple input formats possible!
- Modularized objective
- Modularized output

```
minizinc default.mpc instance.dzn
```

- Other configurations can derive from this!

Using sub-configurations

Dedicated configuration for EC2 service

```
{
  "model": [
    "model/input/gateway.mzn"
  ],
  "intermediate": true,
  "json-stream": true,
  "statistics": true
}
```

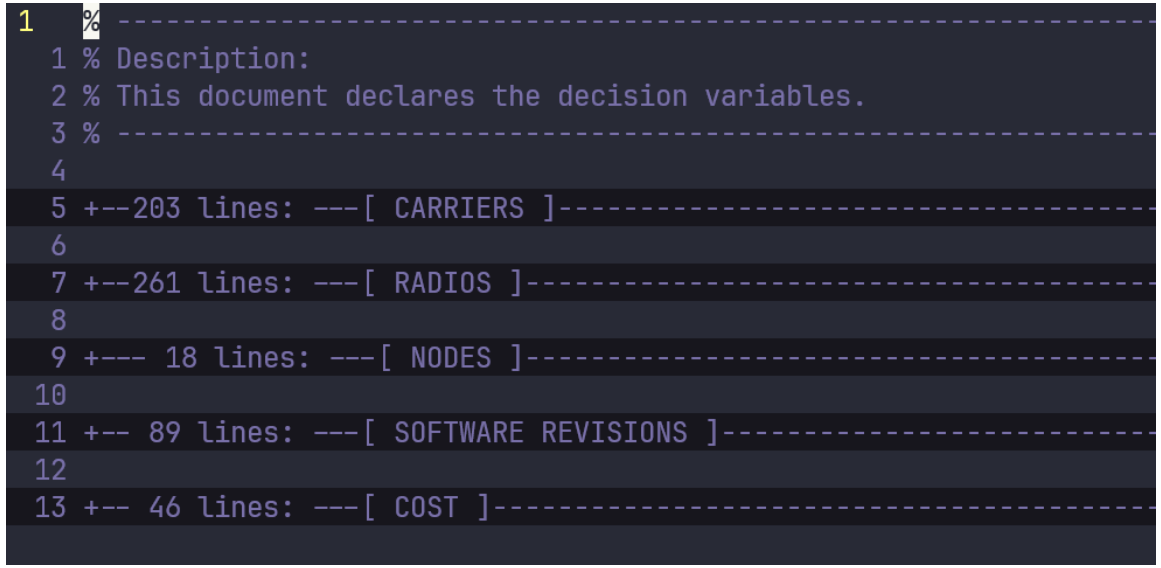
- The *gateway* input file is a *"function"* to the *default* input file

input/gateway.mzn → input/default.mzn

- Can activate additional flags (Json output)
- Allows non-breaking input updates!

```
minizinc default.mpc gateway.mpc instance.dzn
```

Decision Variables



- 109 decision variable declarations
- Categorized by RAN component

Constraints — Core

```
 5
 6 include "all_different.mzn";
 7 include "all_equal.mzn";
 8 include "decreasing.mzn";
 9 include "nvalue_fn.mzn";
10 include "table.mzn";
11 include "value_precede_chain.mzn";
12
13 +--259 lines: ----[ CARRIERS ]----------------------------------
14
15 +--201 lines: ----[ RBBS ]-----------------------------------
16
17 +--415 lines: ----[ RADIOS ]----------------------------------
18
19 +-- 78 lines: ----[ BAND AND RANGE LIMIT]---------------------
20
21 +--276 lines: ----[ CPRI ]-----------------------------------
22
23 +--175 lines: ----[ NODES ]----------------------------------
24
25 +-- 11 lines: ----[ DATA STATUS ]-----------------------------
26
27 +-- 22 lines: ----[ SOFTWARE REVISIONS ]----------------------
~
```

- 83 constraints
  - Table is the most common global constraint
- Categorized according to RAN

A typical constraint
2 out of 83

```
1
2 % Total bandwidth of carrier branches per RF port must not exceed the width of
3 % appropriate frequency band.
4 constraint :: "BAND-LIMIT-CLASSIC"
5 forall(r in Radios, p in Ports, fb in FrequencyBands) (
6   % Downlink
7   sum( b in TxBranches where bandOfBranch(b) == fb )
8       ( (cb_radio[b] == r) * (cb_r_port[b] == p)
9         * freqBwOfBranch(b) )
10   <= freqBwDL(fb)
11   /\
12   % Uplink
13   sum( b in RxBranches where bandOfBranch(b) == fb )
14       ( (cb_radio[b] == r) * (cb_r_port[b] == p)
15         * freqBwOfBranch(b) )
16   <= freqBwUL(fb)
17  );
18
19 % Total bandwidth of AAS carriers per radio must not exceed the width of
20 % appropriate frequency band.
21 constraint :: "BAND-LIMIT-AAS"
22 forall(r in Radios, fb in FrequencyBands) (
23   sum( i in AasIndex, c = AasCarriers[i] where c.frequency_band == fb )
24      ( ( aas_radio[i] == r)
25        * c.fq_bandwidth )
26   <= freqBwDL(fb)
27 );
28
```

- Constraints are annotated (flatzinc, findMus)
- Multiply with bool var, avoids reification

Constrains — Improving

- 24 improving constraints
- Mainly implied/symmetry breaking
  - Fixing dummy values etc.

```
1   % ---------------------------------------
  1 % Description:
  2 % This document implements the constraint
  3 % constraints include implied (redundant)
  4 % constraints, and dominance constraints.
  5 % ---------------------------------------
  6
  7 include "all_equal.mzn";
  8 include "decreasing.mzn";
  9 include "increasing.mzn";
 10 include "value_precede_chain.mzn";
 11
 12 +--180 lines: --- [ IMPLIED ] ----------
 13
 14 +--123 lines: --- [ SYMMETRY BREAKING ] -
```

Optimization Function

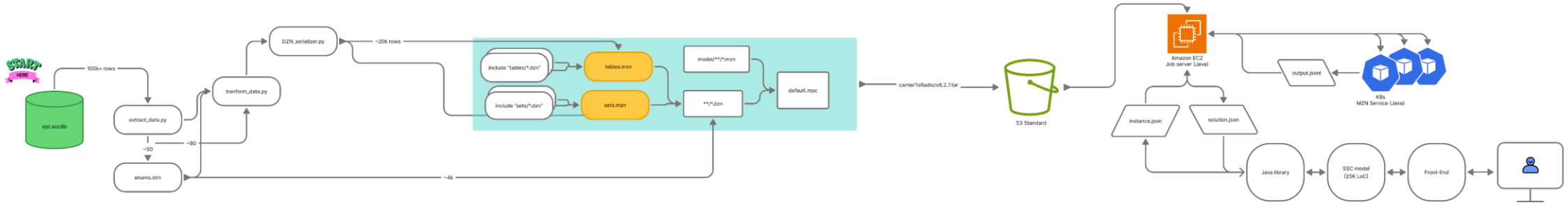
```
1   %  ------------------------------------------------------------
  1 % Description:
  2 % This declares the optimization of the model.
  3 % ------------------------------------------------------------
  4
  5 +-- 67 lines: ---[ Cost assignments ]---------------------
  6
  7 +--173 lines: ---[ Cost function ]------------------------
  8
  9 +-- 58 lines: ---[ Debug output ]--------------------------
```

- Lexicographic Optimization
- Different Optimization Scenarios
  - Prioritize less weight, power etc.

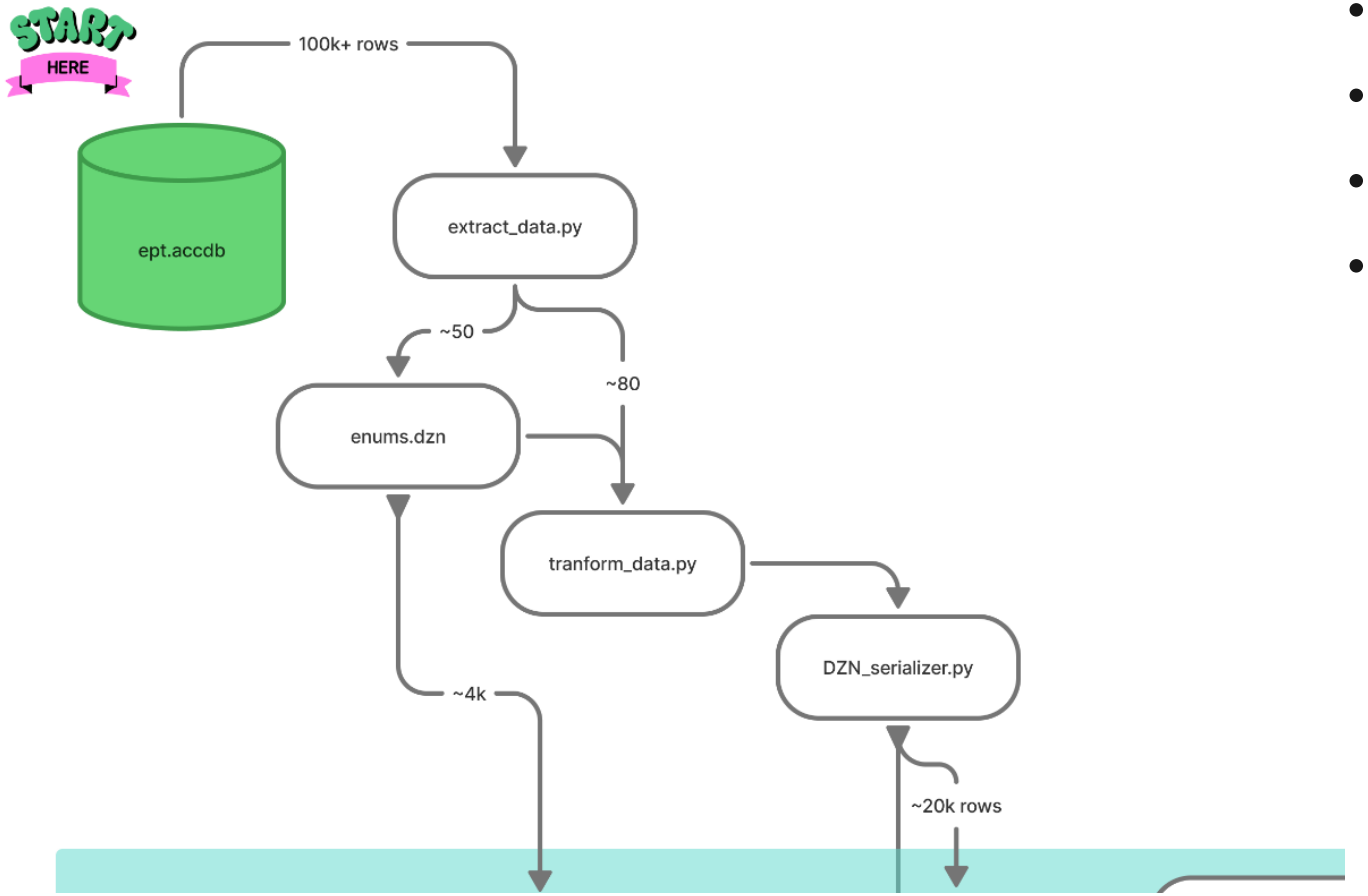
```
27  %    cost   | used | Radio hardware:
  1 %  540360 |      | (freq_ranges: 1, id: XXXXXXXXXXXXX, num_rf_ports: 2, power: 0, volume: 9, weight: 84)
  2 +-- 14 lines: %  540361 |      | (freq_ranges: 1, id: XXXXXXXXXXXXX, num_rf_ports: 2, power: 1, volume: 999, weight
  3 %  561976 |      | (freq_ranges: 1, id: XXXXXXXXXXXXX, num_rf_ports: 4, power: 16, volume: 108, weight: 500)
  4 %  562322 |      | (freq_ranges: 2, id: XXXXXXXXXXXXX, num_rf_ports: 4, power: 2, volume: 5, weight: 48)
  5 %  562336 |  ✅  | (freq_ranges: 2, id: XXXXXXXXXXXXX, num_rf_ports: 4, power: 16, volume: 17, weight: 170)
  6 %  562336 |      | (freq_ranges: 2, id: XXXXXXXXXXXXX, num_rf_ports: 4, power: 16, volume: 17, weight: 170)
  7 +-- 20 lines: %  562348 |      | (freq_ranges: 2, id: XXXXXXXXXXXXX, num_rf_ports: 4, power: 28, volume: 999, weigh
  8 %  606300 |      | (freq_ranges: 4, id: XXXXXXXXXXXXX, num_rf_ports: 8, power: 60, volume: 43, weight: 440)
  9 % ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

Executing our CP model

From data to user

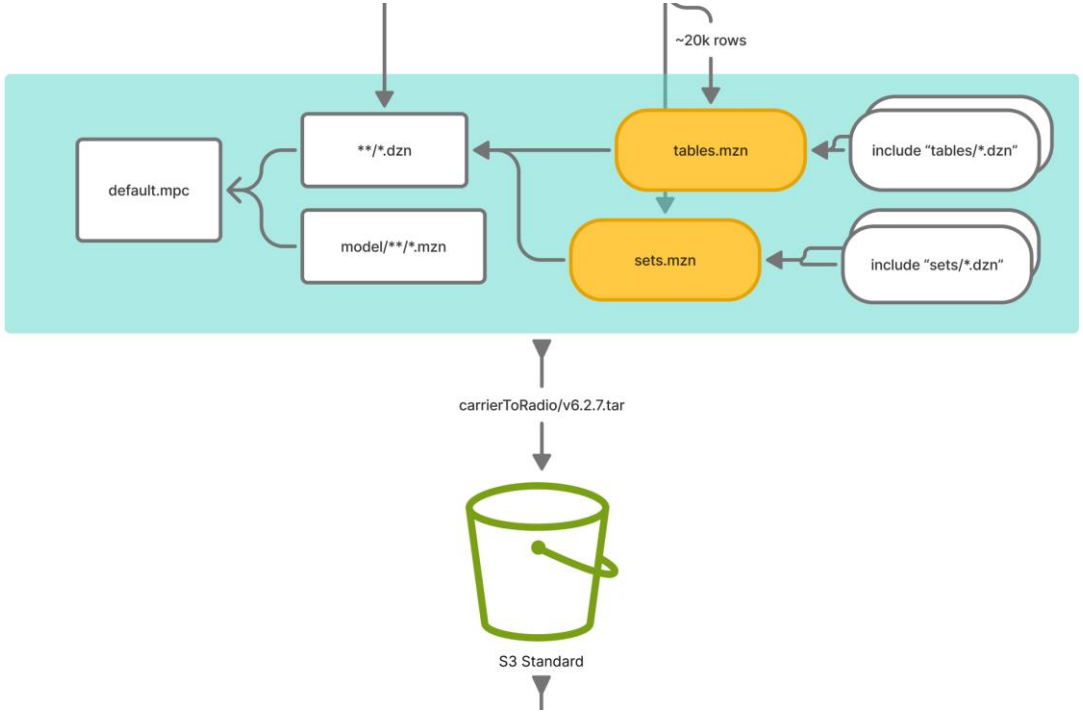


ETL — The Data Pipeline



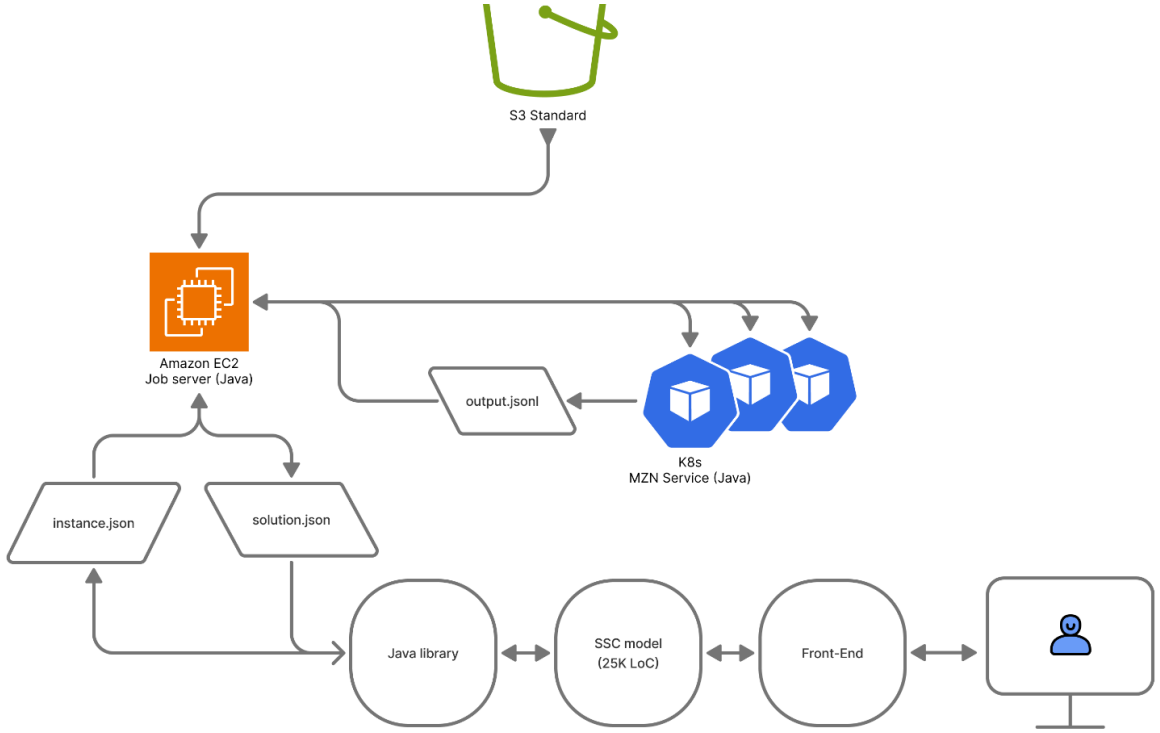
- ~8K python LOC
- Largest enum, 118 chars
- Largest table, ~20k rows
- Challenges:
  - Enums are in the global namespace
  - Special characters in unquoted enums
  - Creating “Null” enums
  - Serializing complex DZN types

Packaging — The model artifact



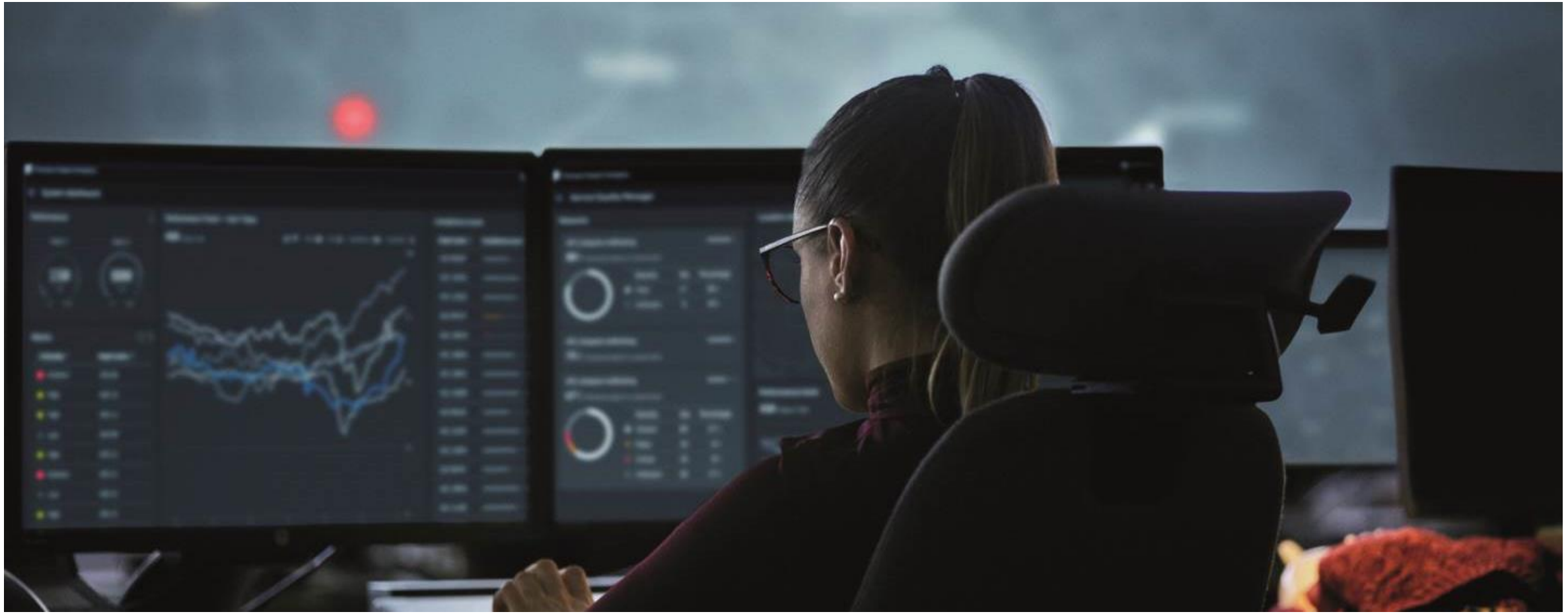
- Simple TAR archive (*.mzn + *.dzn)
- Tagged in Git with SemVer
- Name = model/Semver
- Regression/Integration tested
  - Uploaded to AWS S3 for distribution

Execution — The runtime

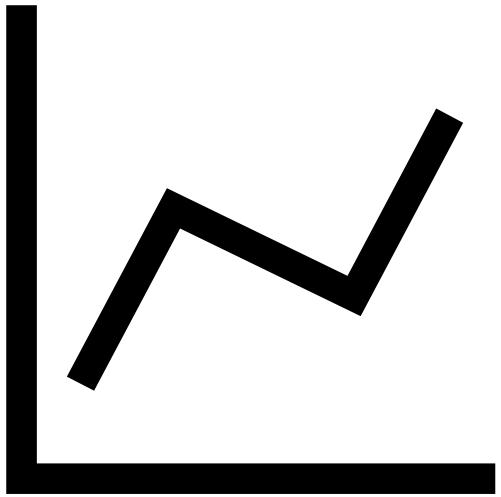


- Continuous delivery
  - SemVer (always get newest)
  - Tiered environments (dev, test, prod)
- Scalability
  - Job queue
  - Parallel processing (K8 cluster)

Maintaining our CP model



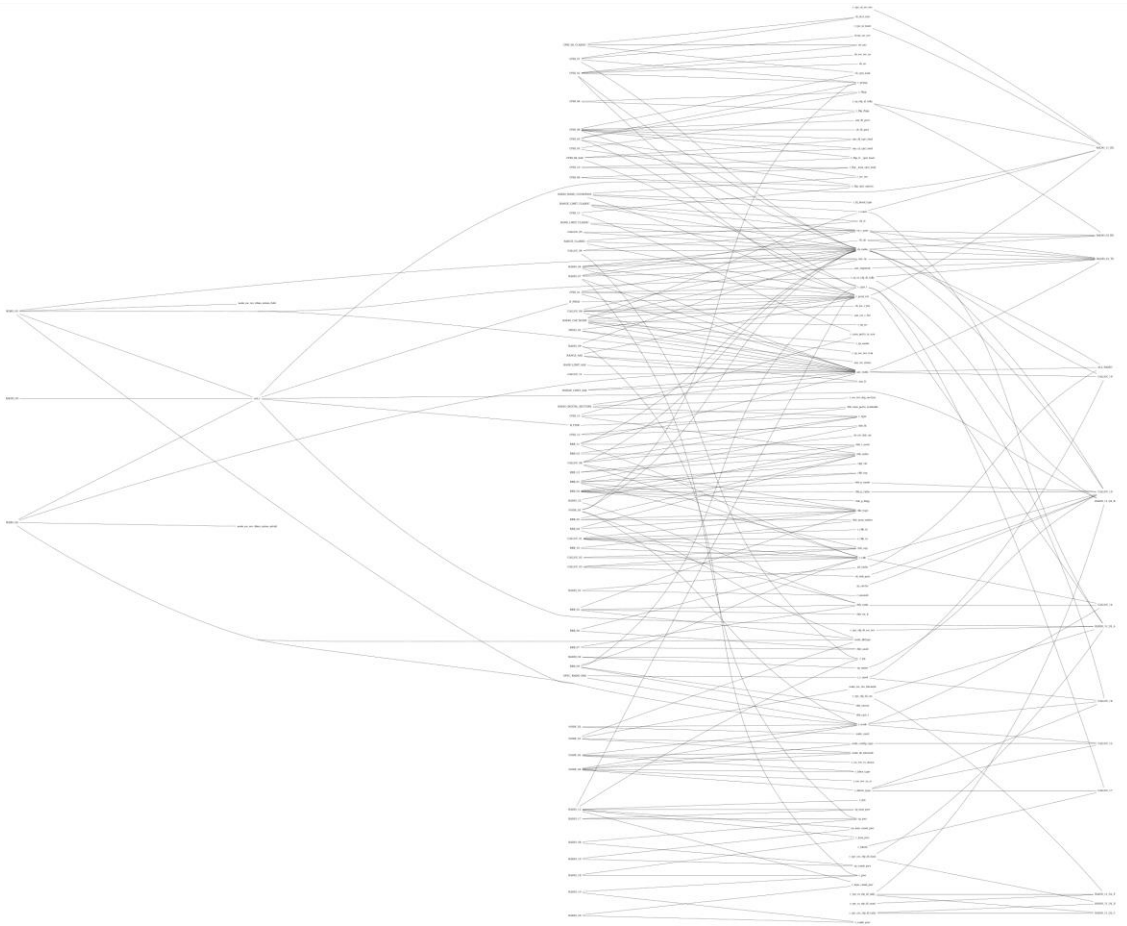
Ever increasing complexity
2x over the last 2 years



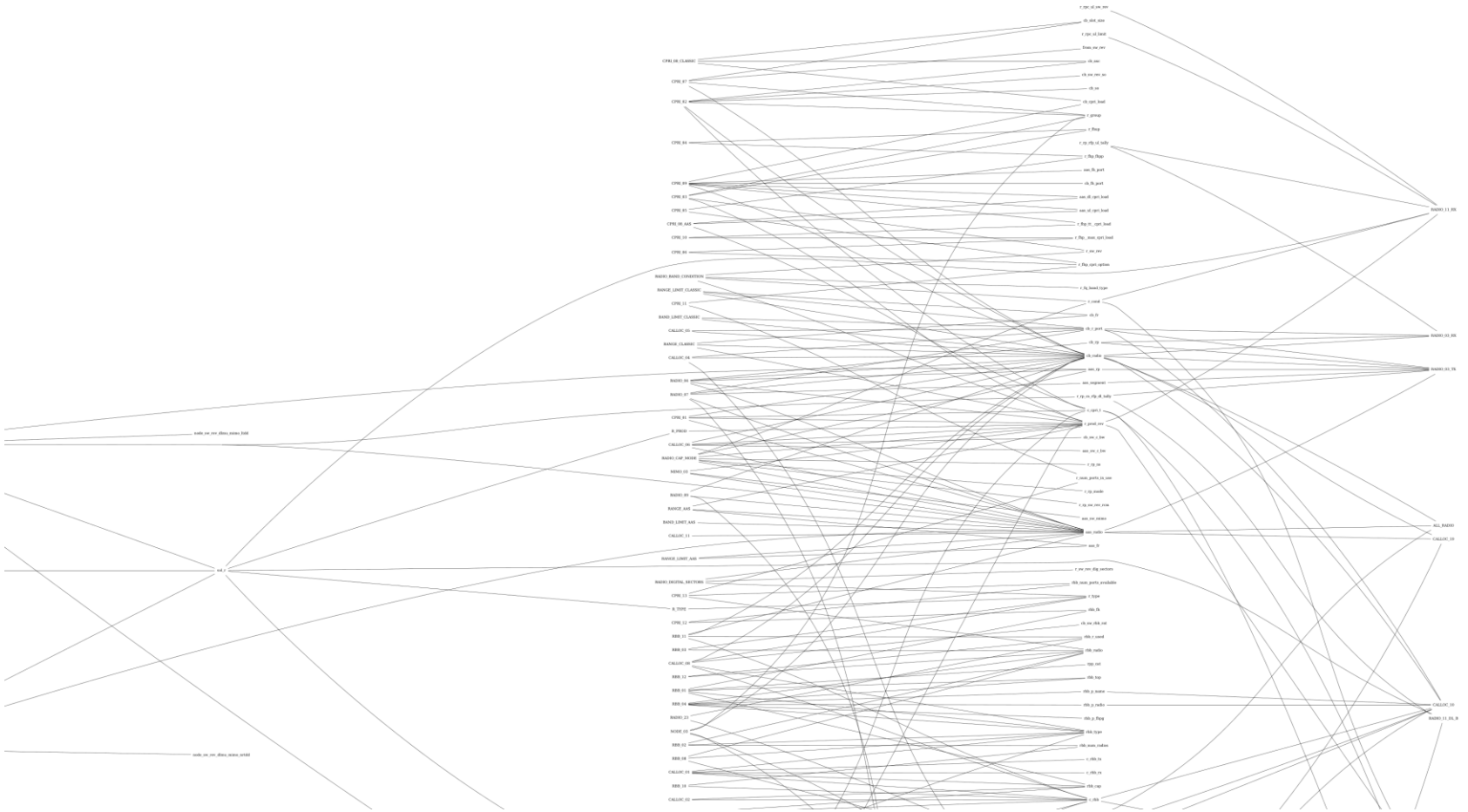
- +4K LoC *.py → +8K LoC *.py
- +4K LoC *.mzn → +8K LoC * .mzn
- +7K LoC *.rst → +10K LoC *.rst
- +60K LoC *.dzn → +120K LoC *.dzn
- 30 testcases → ~140 testcases

- New products
- New rules
- New language features

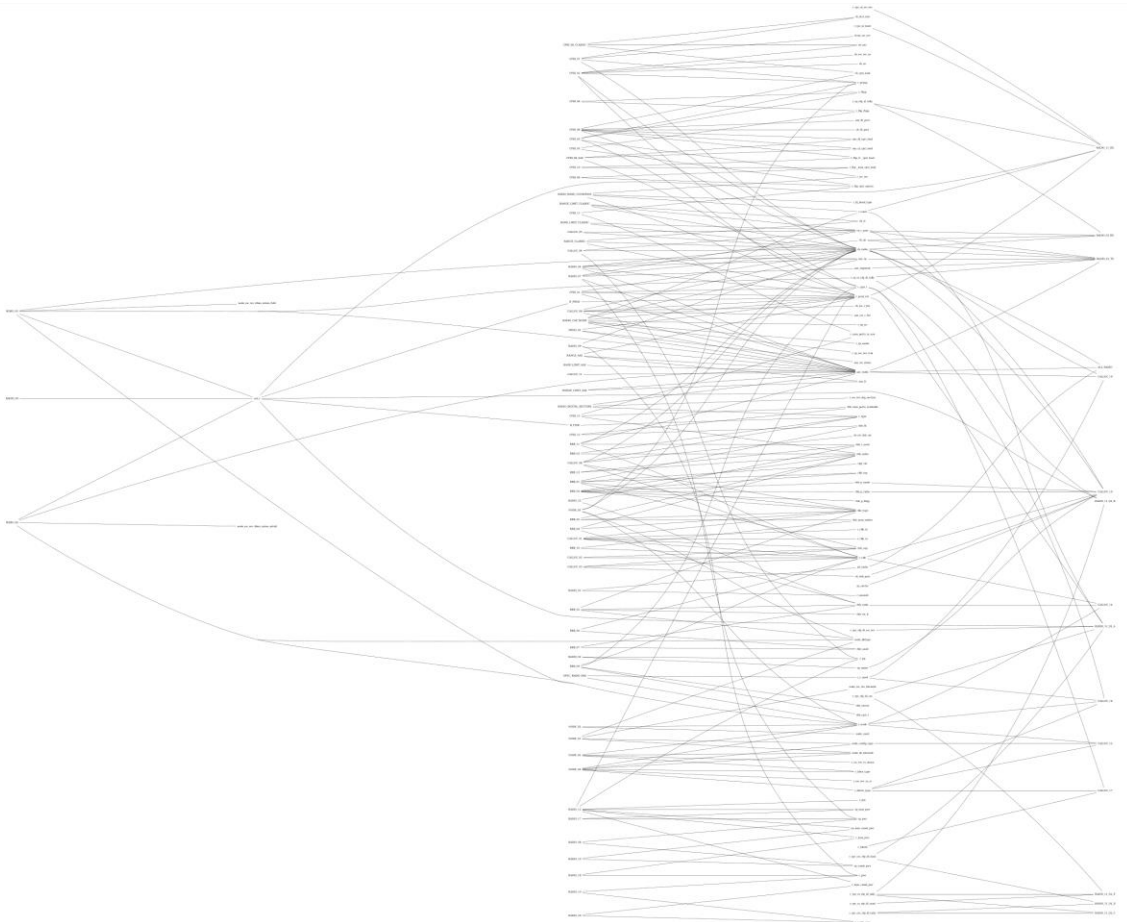
All constraints & decision variables



All constraints & decision variables

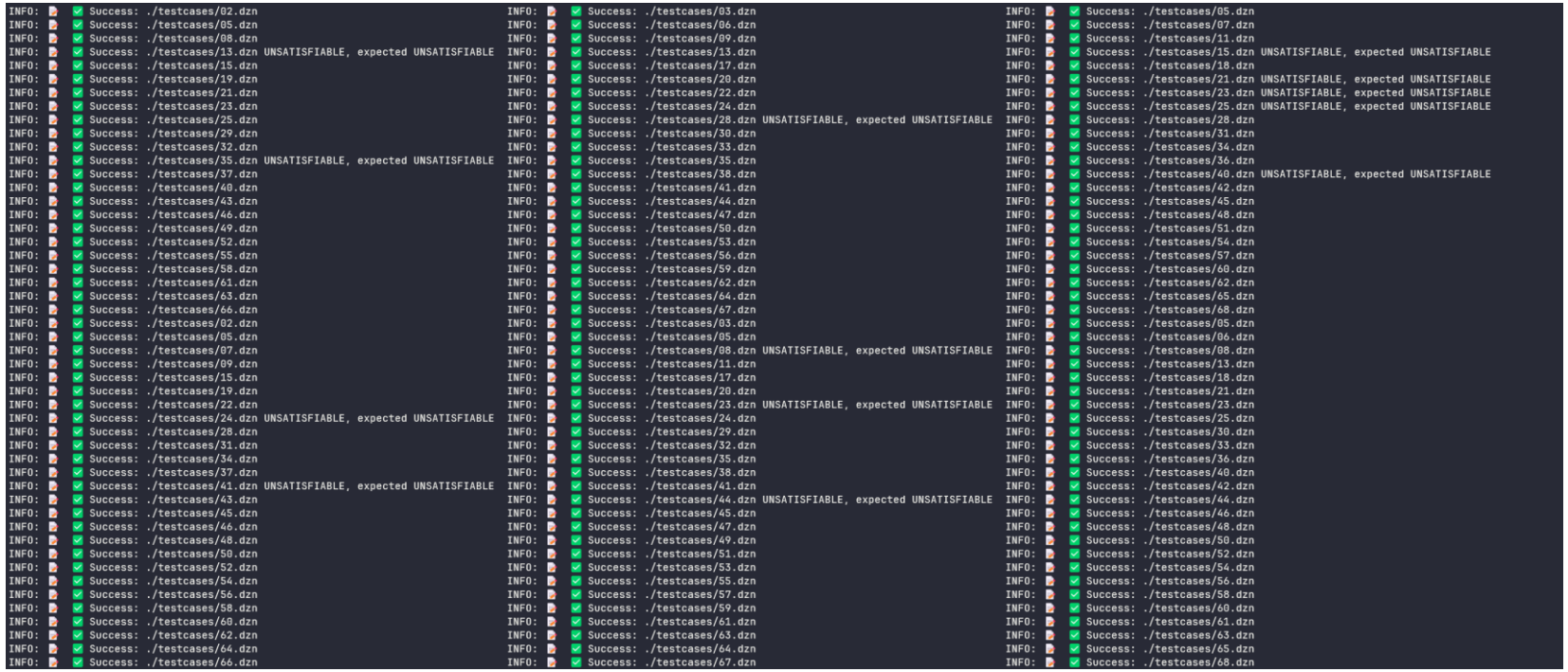


How do we deal with this complexity?



More test

Better test coverage, less regressions & bugs 👏 👏 👏



- Testing both SAT/UNSAT
- New rules
  ➡️ new test
- New bug/regression
  ➡️ new test

Stronger types

Records instead of 2d array of int 👏 👏 👏

```
 5 constraint :: "NODE-01"
 6 forall (n in NodeSet)
 7 ( table( [ node_config_type[n]
 8           , node_bbtype[n]
 9           , node_sw_rev_bbcomb[n]
10           , node_ds_bbcomb[n] ]
11         , NodeConfigTypeBbTypeSubTabWithNull
12         )
13 );
14 %%%
```

```
 5 constraint :: "NODE-01"
 6 forall (n in NodeSet)
 7 ( ( nct : node_config_type[n]
 8   , bbt : node_bbtype[n]
 9   , swt : node_sw_rev_bbcomb[n]
10   , dst : node_ds_bbcomb[n] )
11   r in NodeConfigTypeBbTypeSubTabWithNull
12 );
13 %%%
14
```

- Explicit declaration (no documentation rot)
- No accidental column mismatch
- No accidental type coercions

Records for input/output

Safe & robust object serialization 👏👏👏

```
 5   var SolutionRadioId:          radio_id,
 6   var RfPortNameOrNull:         rf_port_name,
 7   var FrontHaulPortGroupOrNull: fh_port_group,
 8 );
 9 +-- 36 lines: type node_ot = record(----------------------------------------
10 +-- 50 lines: Functions:---------------------------------------------------
11 output :: "gateway_json"
12 +----  8 lines: ---------------------------------------------------------------
13 let {
14 +----  6 lines: ---------------------------------------------------------------
15   array[BranchSet] of cb_ot: carrier_branches =
16       [(
17           branch_id:       cb,
18           type_of:         if cb in TxBranchSet then TX else RX endif,
19           carrier_id:      to_enum(CarrierId, Carriers[c].carrier_id),
20           radio_id:        to_enum(SolutionRadioId,cb_radio[cb]),
21           rf_port_name:    rp_name[cb_radio[cb], cb_r_port[cb]],
22           fh_port_group:   rbb_p_fhpg[ c_rbb[c] , cb_rbb_port[cb] ],
23       ) | cb in BranchSet, c=carrierIdOfBranch(cb) ];
24 +----  7 lines: array[AasIndexSet] of aas_ot: aas =---------------------------
25 +---- 57 lines: ---------------------------------------------------------------
26 % DEFINES: output
27 [
28 +----  8 lines: "{\n",--------------------------------------------------------
29 "  \"CarrierBranch\": " ++
30   if empty(carrier_branches) then "[]"
31   else "[\n" ++ join(",\n", [
32 "     " ++ showJSON(cb) | cb in carrier_branches ]) ++ "\n" ++
33 "  ]"
```

```
19 +-- 49 lines: % ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 1 {
 2   "AasSegment":    [],
 3
 4   "CarrierBranch": [
 5     { "branch_id": 1,
 6       "carrier_id": { "e": "C1" },
 7       "fh_port_group": { "e": "FHPG_UNIT" },
 8       "radio_id": { "e": "R1" },
 9       "rf_port_name": { "e": "RFPORT_C" },
10       "type_of": { "e": "TX" }
11     },
12 +-- 39 lines: {"branch_id": 2, "carrier_id": {"
13 }
14 ----------
15 ==========
16 +---  8 lines: %%%mzn-stat: objective=562347---
```

Exterminating bugs!

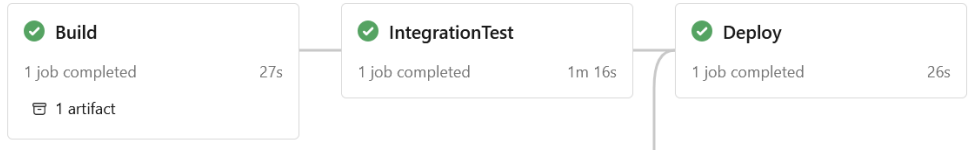
- Compiler errors
  - Stronger types (enums, records etc.)
- Consistency errors
  - Automatic test coverage (~120 testcases) + git bisect
  - Oracle model (SAP SSC)
- Interface errors
  - Strong contracts (types/SemVer)
  - Integration testing
- Solver bugs
  - Compare solvers
- Compiler bugs
  - Segfault/Strange behavior
    - Divide & conquer

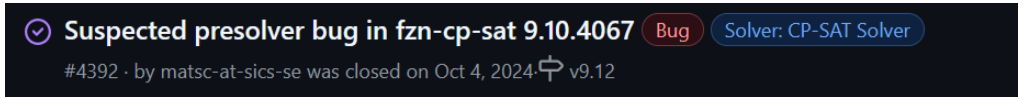
```
.PHONY: check
check: .check
.check: $(MODEL_FILES) testcases/test*.mzn
	minizinc --model-check-only default.mpc
	touch $@
```

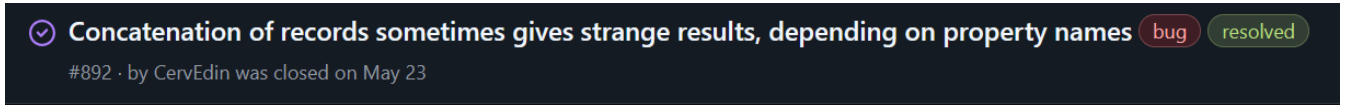
```
.PHONY: test
test:
	MZN_ARGS="$(MZN_ARGS)" ../scripts/test-minizinc.sh $(TESTOPTS) $(MODEL) $(TESTCASES)
```

4.2.1

MAJOR *Minor* patch

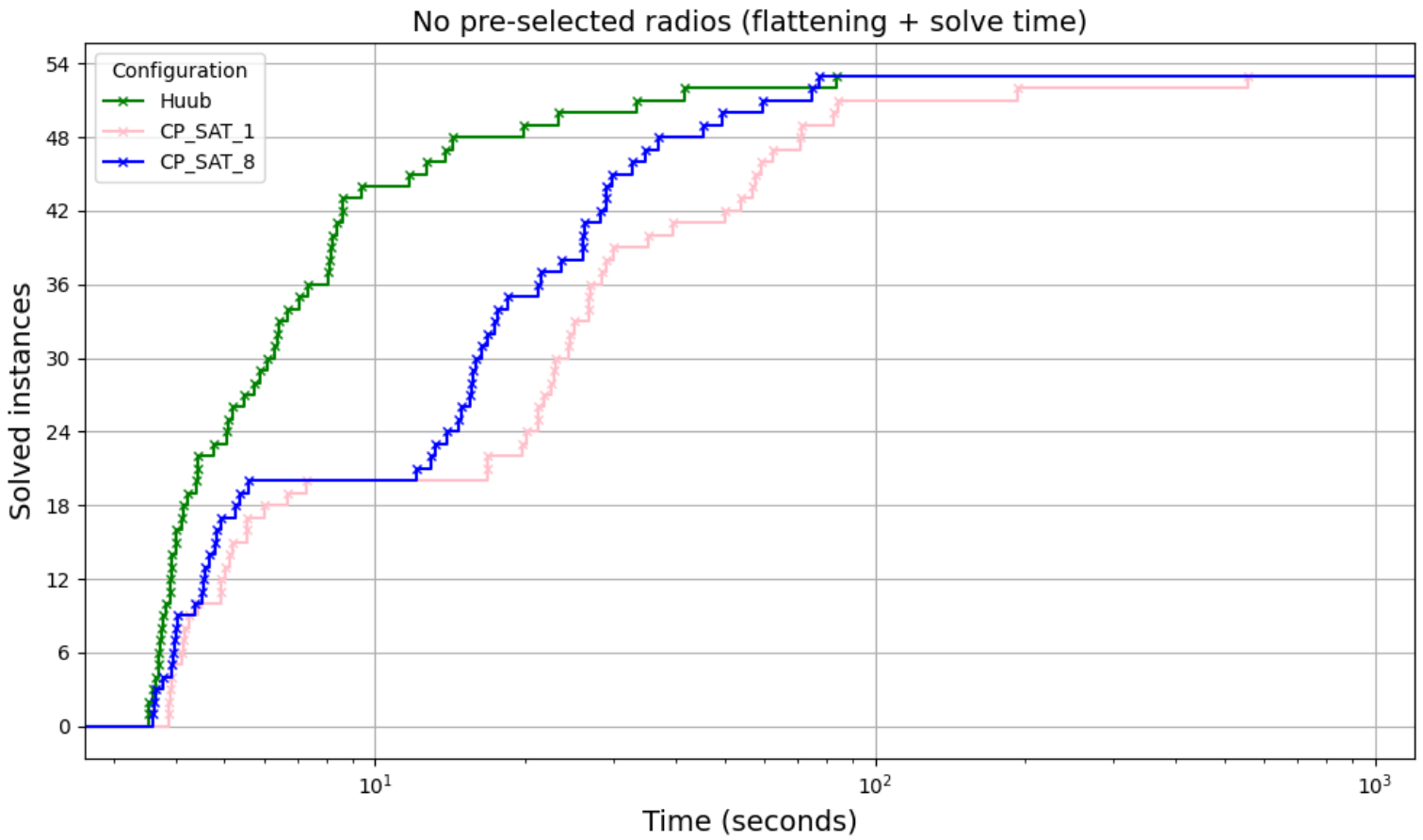






Improving performance

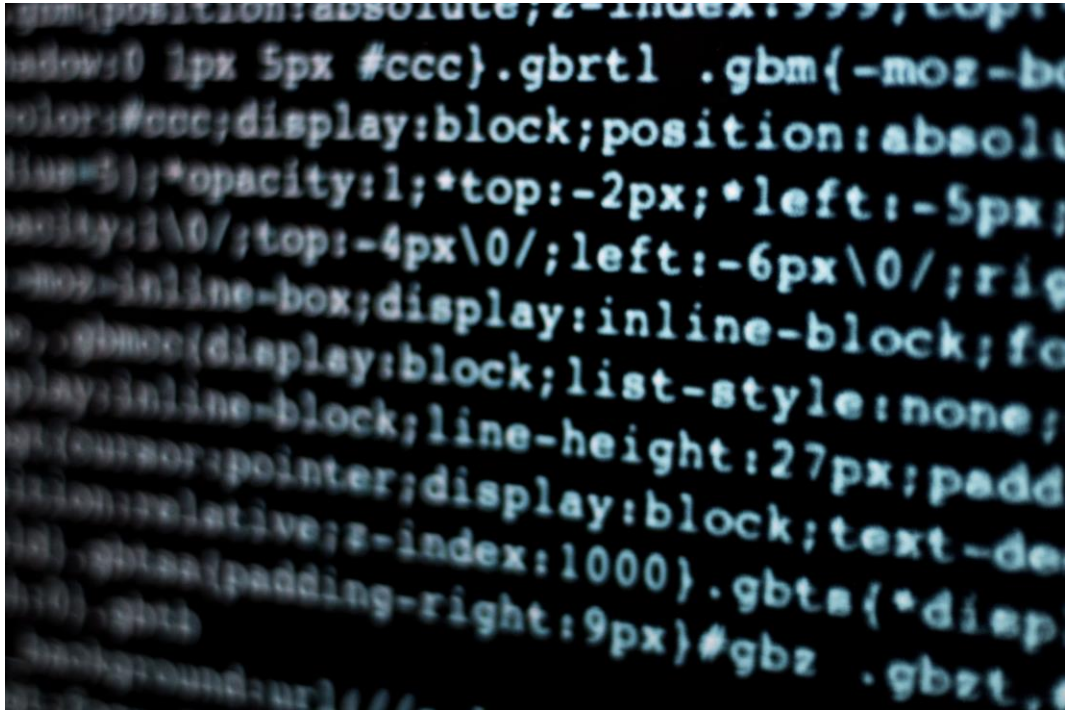
Experimenting with other solvers



Challenges – Pre-filtering

Less code, more constraints

- Large models, ~8k+ LoC of MiniZinc
  - 1/4th is constraints
  - A substantial part is “pre-filtering” or “massaging”
    - Challenging in a DSL
    - Better data-types (caching, indexing etc.)
    - Less complexity, better debugging



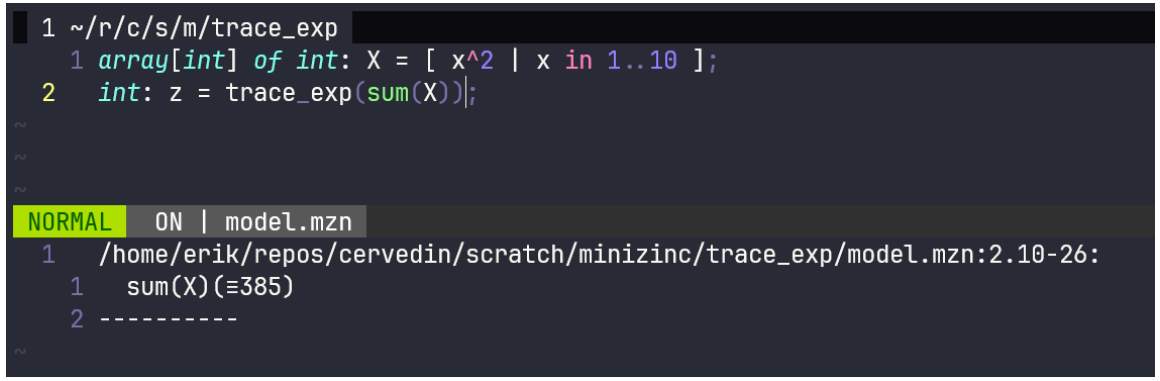
Challenges – Explainability

- Users want explanations, not just **no**
  - Explanations also help debugging
- Soft constraints
  - Challenging for CP
  - Suited for preferences, not explaining
- FindMUS
  - Expensive
  - Cryptic
  - Can be MANY

```
=====UNSATISFIABLE=====
```

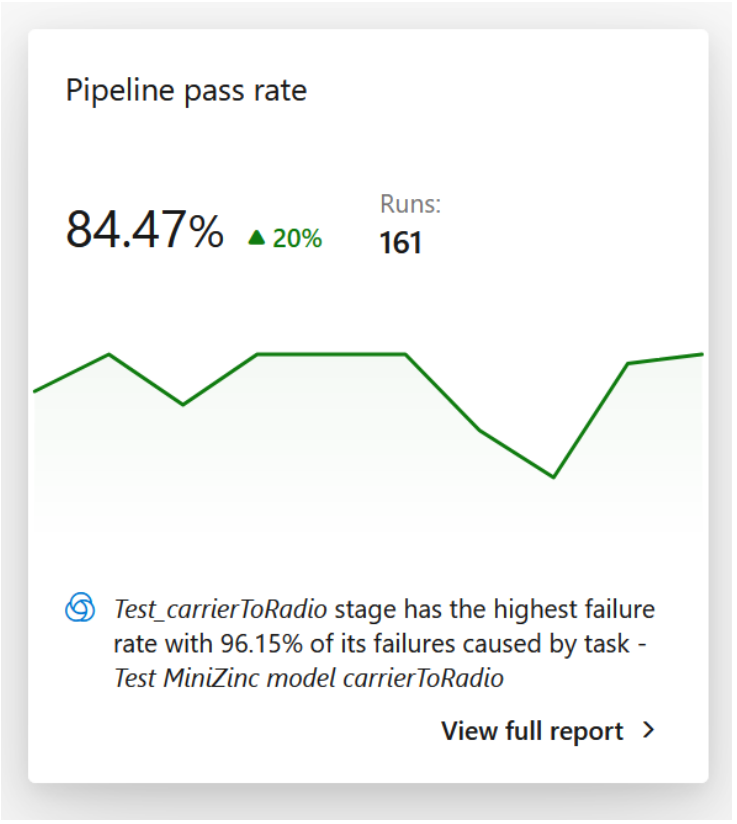
Challenges – Debugging

- Conditional debug output
- `trace_exp`
  - Prints an expression and value
- Black box
  - Poke the box and see what happens!
    - Manual assignments
    - Manually bisect constraints (delete/comment) until UNSAT becomes SAT



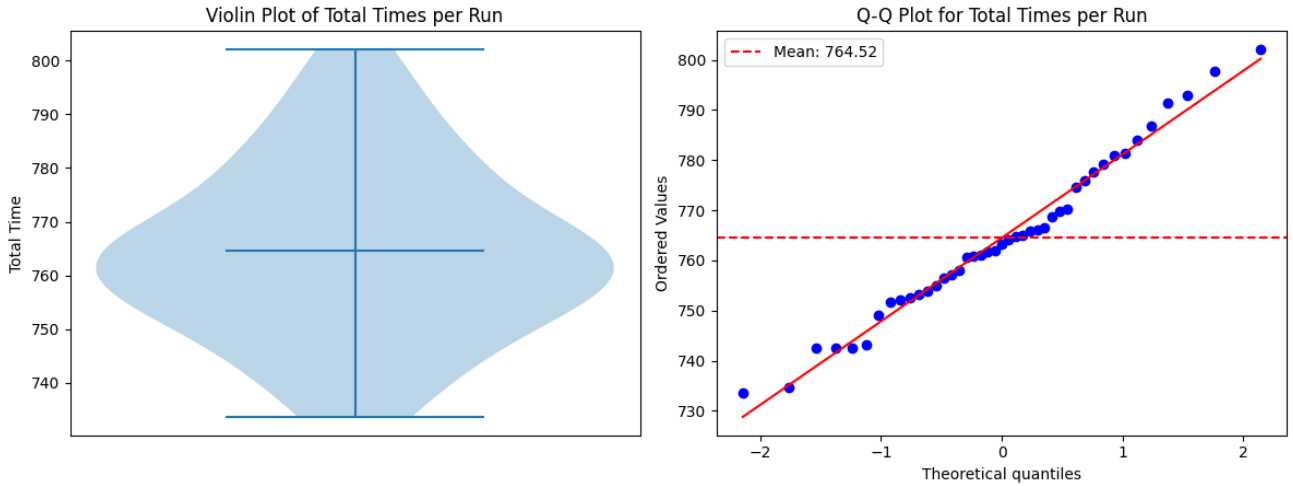
Challenges – Testing

- Hard to test specific parts of the model
  - Need to test the whole model
  - Test using partial assignments
  - Test both for SAT/UNSAT
  - Indirect tests
- Rigorous testing
  - Automated testing (CI/CD/MLOPS)
  - Improving coverage (removing constraint should fail)
  - Large changes difficult (testcase might be symmetrical)



Challenges – Benchmarking Performance

- Small impact, hard to measure
- Parallel solving → high variance
  - Times are normally distributed
  - Benchmark ~5 times
  - Compare flatzinc statistics
- Instance dependent variance
  - Benchmark several different instances



Conclusion

How to build a large CP model & live to tell the tale

- Test, test & test
- Stronger types 👍
- Make incremental changes
- Flexibility & reliability before performance
  - It's easier to make a **correct** program fast than a **fast** program correct.
- Have fun!

Q&A

Acknowledgements

- Mats Carlsson 👩‍🚀
- Co-Workers
  - Fredrik, Sverker, Marko, Samuel, Marko, Zdravko, José, Mariia, Nils, Olle, Danyal
- MiniZinc team (Jip, Guido, Jason, Peter & friends)
- OR-Tools team (Laurent & friends)
- GeCode (Christian, Guido again, Mikael)
- Uppsala Optimization group (Mats, Pierre, Justin, Maria, Ramiz & friends)


Imagine Possible
ericsson.com/careers