# **Formal Verification of Distributed Network Control Planes**

- Aarti Gupta **Princeton University**
- Ryan Beckett (Microsoft Research), Ratul Mahajan (Univ. of Washington),
  - Divya Raghunathan (Princeton), Tim Alberdingk Thijm (Princeton), and David Walker (Princeton)



Joint work with:



### **Protocol + Configuration Control Plane Protocols RIP, OSPF, BGP, ...**

7.2.0.*	port	5
9.2.*.*	port	3

### Forwarding Tables

### **Runtime Behavior**



### **Network Control Plane**

Primary goal is to get traffic from point A to point B but ...



**Configuration Vendor-specific Thousands of lines** 

**Distributed program Assembly-language** 

### **Misconfiguration is a BIG problem**

#### BGP errors are to blame for Monday's Twitter outage, not DDoS attacks

No, your toaster didn't kill Twitter, an engineer did



#### **Unions want Southwest CEO removed after IT** outage

Massive route leak causes Internet slowdown

Posted by Andree Toonk – June 12, 2015 – BGP instability – No Comment

Home / Cisco Security / Security Advisories and Alerts

Internet Services

### Motivated many formal verification efforts

### • Data plane verification checks a *snapshot* of the network

- Analyzes the given forwarding rules at routers to check various properties
- Based on model checking, symbolic simulation, SAT/SMT/BDD techniques
- Routinely applied in large data centers (~10k routers)

#### **Data Plane Verifiers**

. . .

Anteater	[Mai 2011]
HSA	[Kazemian 2012]
Veriflow	[Kurshid 2013]
NoD	[Lopes 2015]
Symmetries	[Plotkin 2016]

t routers to check various properties imulation, SAT/SMT/BDD techniques 's (~10k routers)

### **Motivated many formal verification efforts**

- Control plane verification checks router configurations that determine the forwarding rules
  - Verifies all possible data planes that result from the configurations

  - So far, they have been demonstrated on 2-3 k routers (max)

#### **Control Plane Simulators**

C-BGP [Quotin 2005] Batfish [Fogel 2015]

- Bagpi
- ARC
- ERA
- FastP
- Plankt
- Hoyan

#### **Our** w

• Based on model checking, graph-based techniques, SAT/SMT/BDD techniques

#### **Control Plane Verifiers**

ipe	[Weitz 2016]	symbolic execut
	[Gember-Jacobsen 2016]	graph-based
	[Fayaz 2017]	
Plane	[Lopes 2019]	semi-symbolic
ton	[Prabhu 2020]	oonn oynnoono
n	[Ye 2020]	
vork	[Becket 2017, 2018,]	fully-symbolic
	6	



### **Network Properties**



reachability



Router or subnet equivalence





#### no transit







#### no black holes



# **Modeling Network Control Planes** in MineSweeper

### Ryan Beckett, Aarti Gupta, Ratul Mahajan, David Walker:

A General Approach to Network Configuration Verification. SIGCOMM 2017: 155-168



## **A Generic Routing Protocol**



Idealized RIP: A simple routing protocol The origin creates an *initial announcement* stating it has a path to destination d



[Griffin and Sobrinho: Metarouting]

**Other nodes that receive the announcement** pass it on to their neighbors, *possibly* modifying it

When nodes receive multiple announcements, they choose a best one

**Eventually (hopefully), the system converges** to a stable solution: all nodes are happy



# **Stable Paths Problem (SPP)**

#### Imperative, Stateful Program

```
process spvp(u)
begin
      receive P from w \rightarrow 
            begin
                  \operatorname{rib-in}(u \Leftarrow w) := P
                  if rib(u) \neq best(u) then
                  begin
                        \operatorname{rib}(u) := \operatorname{best}(u)
                        for each v \in peers(u) do
                        begin
                               send rib(u) to v
                         end
                  end
            end
end
```

- Prior work on reasoning about protocol/network convergence
- We apply it for verifying *network configurations of protocols*

#### [Griffin et al. 2002]

#### Logic Model

Choices  $(P, u) = \dots$ Best  $(u) = \dots$ 

#### Each node is locally stable





### Minesweeper: Insights

Network protocols are designed to generate <u>stable</u> paths *i.e., routers exchange messages to make <u>best</u> choice, which stays stable* 

#### **Our Idea**

Capture network control plane behavior in terms of logical constraints, such that satisfying solutions are stable paths in the network

**Analogy to Program Verification** Program: Satisfying solution represents a path in the program graph Network: Satisfying solution represents *stable* paths in the network graph

But: arbitrary (not well-structured) graphs in network topology, a single solution corresponds to many paths; we target a *stable routing tree* 



11

### **Minesweeper: Key Choices**

Choice 1: Model routing <u>graphs</u>, not paths at a time

- Too many paths, but all paths share the same graph
- In the data plane, reasoning is done per-packet because there is no interference between packets along different paths
- But, in the control plane, routing messages along different paths *interact* with each other – modeling this interaction can be expensive!

perform *search* on final stable states

- Familiar lesson from symbolic model checking vs. bounded model checking
- Solve a search problem over state space, use modern SAT/SMT solvers
- Often scales better than *computing* sets of states iteratively

Choice 2: Don't compute states due to exchange of routing messages, but



### **Minesweeper Approach**

- **SRP** (Stable Routing Paths) a general, logical control plane model Framed in terms of **local** route processing constraints at a node
- Applies translation to **SMT-based logic** for **verification**
- Heavily **optimized** to make the resulting tool practical





### Minesweeper: SRP Model

Topology:	$\mathbf{G}=(\mathbf{V},$
Attributes:	$A_{\perp} = A$
Preference relation:	$\prec$ : A ×
Transfer function:	transfei
Initial value:	$a_d \in A_\perp$

An SRP solution is a labeling (based on neighbors): L:  $V \rightarrow A_{\perp}$ An SRP solution is locally stable, i.e., each node is happy

Each SRP solution corresponds to a forwarding relation





### **SRP Models for Popular Protocols**

- Many network protocols are used in practice
  - RIP
  - OSPF
  - BGP (eBGP, iBGP)
  - Static Routing
- Minesweeper handles them all as SRPs

uniform model allows handling fancy features like route redistribution etc.



### **Example SRP**

### **Routing Information Protocol (RIP)**



#### **Attributes:**

 $A = \{0..15\}$ 

#### **Preference relation:**

 $a \prec b \Leftrightarrow a < b$ 

#### **Transfer function:**

transfer(e, a) =  $\begin{cases} \bot & \text{If } a=15\\ a+1 & \text{otherwise} \end{cases}$ 



### **SMT Encoding for Verification**



Does P hold in the network?

# Network Encoding (SRP): N

Network Property (Negated): ¬P

Satisfiable: Property violation

Unsatisfiable: Property holds for all data planes



### **Example: Reachability Property**



Can router R1 reach host H1?

 $canReach_{R3} \leftrightarrow forwards_{R3,H1}$ 

 $canReach_{R2} \leftrightarrow$  $(forwards_{R2,R3} \land canReach_{R3}) \lor$  $(forwards_{R2,R1} \land canReach_{R1}) \lor$  $(forwards_{R2,N1} \land canReach_{N1})$ 

 $canReach_{R1} \leftrightarrow forwards_{R1,R2} \wedge canReach_{R2}$ 

**Property:** canReach<sub>R1</sub>



### **Encoding Transfer Function**

Attributes are like states





SMT theories: bit vectors, LIA



### **Common Network Design Features**

Features	Implemente
<b>OSPF Intra-area</b>	$\checkmark$
<b>OSPF</b> Inter-area	
eBGP Local-pref	
eBGP Communities	
eBGP MEDs	$\checkmark$
eBGP Path Prependin	ig 🗸
eBGP Aggregation	$\checkmark$

#### ed

Cor	ntinued	

iBGP	
<b>Route Reflectors</b>	
Static Routes	
<b>Route Redistribution</b>	
Multipath Routing	$\checkmark$
Access Control Lists	
IPV6	×



### **Properties Supported**

ス Properties

Reachability **Bounded Path Le Equal Path Ler Disjoint Paths Multipath Consist Routing Loop** Black Holes **ECMP** Load Balar **Router Equivale** 



Minesweeper Limitations: Checking multiple destinations is expensive

	Implemented
	$\checkmark$
ength	
ngths	$\checkmark$
IS	
stency	$\checkmark$
OS	
incing	
ence	$\checkmark$

### Does not support convergence, quantitative/probabilistic properties



## **Minesweeper Evaluation**

- Can Minesweeper find real bugs?
   Ran on a collection of 152 legacy networks
   1-23K lines of configuration each
- How well does Minesweeper scale?
   Tested on a collection of synthetic data center benchmarks
   Compared verification time across a wide variety of properties



# **Evaluation: Bug Finding**

- Management interface reachability Found 67 violations of the property
- Local equivalence of routers
  - Found 29 violations
  - Example: ACL has missing entry
- Blackholes occur only at the network edge Found 24 violations of the property
- **Reachability is the same after any 1 failure** Found no violations of the property











### **Evaluation Results**



### Management interface reachability < 60 ms



< 400 ms





### **Evaluation: Scalability**





#### **Black holes only occur** at the network edge < 1.5 sec

#### **Reachability is the same after** any single link failure

< 350 ms









Verification Time (Minesweeper)

# of devices

Other technologies, such as simulation, suffer similar, although less severe trends.

industrial data centers



# Abstraction Part 1

Ryan Beckett, Aarti Gupta, Ratul Mahajan, David Walker: Control plane compression. SIGCOMM 2018: 476-489



### Network Abstraction based on Symmetry



#### <u>Goal</u>: Compute a small *compressed* network with a "similar" solution to the big one





### A pair of abstraction functions: (f, h)

abstracts network topology



### **Formalizing a Compressed SRP**

abstracts route announcements





### **SRP** Abstraction

### A pair of functions: (f, h)





### SRP Abstraction

### A pair of functions: (f, h)









# **Soundness of SRP Compression Abstraction**

**Theorem:** If an **abstraction** satisfies certain requirements (forallexists requirement, transfer equivalence requirement), then it will compute **similar global solutions** as its related concrete network.



 $\forall v, h(L(v)) = L'(f(v))$ 

L is a solution in **concrete** SRP L' is a solution in **abstract** SRP

similar (modulo h) best routes





d  $b_1$  $b_2$  $a_1$  $a_2$ 

# Corollary

- Valid abstractions preserve:
  - (1) Reachability (2) Routing Loops (3) Hop Count (4) Multipath Consistency (5) Waypointing
  - But not fault-tolerance









- The Bonsai algorithm compresses real networks by a factor of 5-7 in the number of nodes and 5-100 in the number of edges.
- It preserves many path properties, such as reachability, but not fault tolerance.
- We have proven it correct for a wide range of routing protocols.

### **Bonsai: Control Plane Compression**





### **Synthetic Benchmarks** [MineSweeper verifying all-pairs reachability]






### BUT ... what if there is no topological symmetry?



## Abstraction Part 2

### Ryan Beckett, Aarti Gupta, Ratul Mahajan, David Walker:

Abstract interpretation of distributed network control planes. POPL 2020



## Leveraging Abstract Interpretation

#### Use abstraction on route announcements to combat complexity

- Formalize theory of abstraction for routing protocols
  - Combine abstract interpretation ...
  - ... with the theory of routing algebras.

[Cousot and Cousot 1977] [Sobrinho 2005]



### **Routing algebra basics**





## **Example: Abstracting (simplified) BGP**



- Routing algebra: (*R*, ⊕, *F*, *0*, ∞)
- Routing messages R: (local\_pref, path, comm)
- Transfer F on edge (i, j): add j to path, local pref update

• **Abstract** routing messages: (R, comm<sup>\*</sup>) R: reachability marker comm\*: (0, 1, \*} for each c





## Abstractions as <u>abstract</u> routing algebras

Concrete Algebra  $A = (S, \bigoplus, F, 0, \infty)$ Abstract Algebra  $A^{\#} = (S^{\#}, \bigoplus^{\#}, F^{\#}, 0^{\#}, \infty^{\#})$ 

	Property	De
(1)	$\oplus, \oplus^{\sharp}$ commutative	$c \in$
(2)	$\oplus$ , $\oplus^{\sharp}$ associative	$c \oplus$
(3)	$\oplus^{\sharp}$ monotone	$a \sqsubseteq$
(4)	$\oplus, \oplus^{\sharp}$ sound for $\alpha$	$\alpha(a)$
(5)	<i>F</i> , $F^{\sharp}$ sound for $\alpha$	$\alpha(j$
(6)	<i>F</i> <sup>♯</sup> monotone	$a \sqsubseteq$

finition

 $\oplus d = d \oplus c$  $\oplus (d \oplus e) = (c \oplus d) \oplus e$  $\sqsubseteq^{\sharp} c \wedge b \sqsubseteq^{\sharp} d \implies a \oplus^{\sharp} b \sqsubset^{\sharp} c \oplus^{\sharp} d$  $(c \oplus d) \sqsubseteq^{\sharp} \alpha(c) \oplus^{\sharp} \alpha(d)$  $f(c)) \sqsubseteq^{\sharp} f^{\sharp}(\alpha(c))$  $\sqsubseteq^{\sharp} b \implies f^{\sharp}(a) \sqsubseteq^{\sharp} f^{\sharp}(b)$ 





#### Soundness theorems

**Theorem 2:** If A<sup>#</sup> is uniquely converging, then for **any** asynchronous

**Theorem 1:** For a fixed asynchronous schedule, an execution of abstract algebra A<sup>#</sup> is sound with respect to an execution of concrete algebra A.

schedule, an execution of abstract algebra  $A^{\#}$  is sound with respect to A.



#### **Abstraction: key benefits**

- Abstraction can (often) give precise answers
  - In our experiments, we got precise answers on 95% of real networks
- Abstraction improves performance
  - Tracking less information -> less memory
  - Smaller state space -> fewer iterations to converge -> less time
  - Creates opportunities for sharing (e.g., for multiple destinations)
- Abstraction can enable new network analyses
  - Example: potential hijacking analysis



## **ShapeShifter: Fast Reachability Analysis**

- Goal: allow fast analysis, while minimizing loss in precision on real networks
- What to abstract?
  - Throw away: AS-path, protocol decision process variables
  - Keep: BGP communities, BGP origin, protocol used
  - Example map [dest  $\mapsto$  abstract route] :
    - $[168/8 \mapsto ([0,*,0], \{R1\}, \{BGP\})]$
    - Means: "second community may be attached or not, on a route originating from R1 in the BGP protocol"
- Other features
  - BDD-based representations for prefix sets for improved sharing
  - Message scheduling heuristics for improving performance
  - Modeling multiple protocols, iBGP, ...



#### **ShapeShifter: Evaluation on real networks**

- - 1-2 orders of magnitude speedup over Batfish (concrete simulator)



Abstract simulation finished in less than 40 sec. on each of the 127 networks





#### ShapeShifter: Synthetic datacenter results



Simulation time vs. data center size for verifying all-pairs connectivity





## Can we scale the general SMT-based approach?

## (Analogy: static program analysis vs. **SMT-based program verification**)

#### **BUT** ...

## Abstraction Part 3

Divya Raghunathan, Ryan Beckett, Aarti Gupta, David Walker: ACORN: Network Control Plane Abstraction using Route Nondeterminism. (Under submission)

#### **Inspiration: Minesweeper's SMT-based approach**



- N : SMT formula representing network behavior • Satisfying assignments of N represent stable paths • P : SMT formula representing property to be checked





#### **Goal:** improve scalability

Two-part Approach

Abstract away route selection while preserving soundness

2. SMT encoding and solvers Symbolic graph-based encoding [Bayless et al. AAAI 2015] Use SMT solvers with specialized graph-theories, e.g., MonoSAT



# **1. Hierarchy of Nondeterministic Routing Choice (NRC) Abstractions**

#### 51

#### **Strategy: Abstract away route selection**

- Insight: Verifying certain properties (e.g., reachability) may not need to identify the best route available
- Modeling route selection is expensive • Especially for complex protocols like BGP
- One difference between data plane and control plane: route selection • Data plane verifiers routinely scale to *several thousands* of routers • Question: can we get closer to their performance?

#### **Key Idea:** Nondeterministic Routing Choice (NRC) Abstractions

- Each router nondeterministically chooses one of the routes received
- Any available route, not necessarily the best, may be chosen
  - Route announcement fields involved only in route selection can be abstracted away
  - Any chosen route must be compliant with policy
    - For example, routes filtered based on community tags will not be chosen
- Abstract network model N' has more routes than real network N • N' includes best routes and other policy-compliant routes

  - N' overapproximates N
  - A hierarchy of NRC Abstractions based on routing fields



#### **SMT-based verification with NRC abstraction**



- N': SMT formula overapproximates network behavior
  - Satisfying assignments of N' represent abstract stable paths
- P : SMT formula representing property to be checked





#### **Motivating Example with BGP**

Real network N



Route announcement from d reaches a  $\Rightarrow$  a can reach d in real network

4 solutions: **a** can reach **d** in <u>all</u> abstract stable routing trees V



Abstract network N'



5	5
$\mathcal{I}$	$\mathcal{I}$

#### Soundness of NRC abstractions

- We use the SRP model of the network control plane
- NRC is formulated in terms of an *abstract* SRP
  - where the preference relation is a partial order, rather than a total order
  - Partial order must be consistent with the concrete total order

- Soundness: Abstract SRP S' overapproximates corresponding SRP S
  - Stable solutions of SRP S are guaranteed to be contained in stable solutions of SRP S'

#### **Goal:** improve scalability

Two-part Approach

Abstract away route selection while preserving soundness

2. SMT encoding and solvers Symbolic graph-based encoding [Bayless et al. AAAI 2015] Use SMT solvers with specialized graph-theories, e.g., MonoSAT



# **1. Hierarchy of Nondeterministic Routing Choice (NRC) Abstractions**

57

#### **MonoSAT solver**

- SMT with graph theory solver
- Uses symbolic graphs, graphs with a Boolean variable per edge  $\bullet$



[Bayless et al. AAAI 2015]

Symbolic graph  $G_{RE} = (G, RE)$ G = (V, E) $RE = \{re_{uv} \mid (u, v) \in E\}$ 

Reachability predicate:  $G_{RE}$ .reaches(d,v)



### Stable paths: symbolic graph-based encoding

- Encode abstract SRP S' with SMT formula N'

Formula F over RE and other variables, can also use graph-based predicates • Satisfying assignment of  $F \rightarrow$  subgraph of G s.t.  $re_{\mu\nu}$  is true

• Key idea: abstract stable routing trees of S' are captured by symbolic graph solutions



## **Benefits of NRC and graph-based encoding**

- Fewer variables

  - Community attribute sufficient for most policies evaluated

#### Expensive transfers can become irrelevant during solver search

- neighbors to compute best
- other neighbors become irrelevant

• Route announcement fields used only in route selection are discarded

Concrete formulation (i.e., no NRC abstraction) considers transfers from all

• In the abstract formulation once a symbolic edge variable is assigned true,

#### ACORN prototype verifier



- Input format is an intermediate representation (IR)
- Two backend SMT solvers: MonoSAT and Z3

esentation (IR) and Z3

#### **ACORN Evaluation**

- 1. Relative performance of NRC abstraction (with / without) 2. Relative performance of graph theory capable SMT solver (MonoSAT / Z3)

Four experiment settings:

- abs mono: NRC abstraction using MonoSAT
- **abs\_z3**: NRC abstraction using Z3
- mono: no abstraction using MonoSAT
- **z3**: no abstraction using Z3



#### **Benchmark examples**

#### 1. Data center examples with FatTree topology (to evaluate scalability)

- Valley-free policy
- Properties: reachability (single-src), valley-free property

#### 2. Wide Area Network (WAN) examples

- Topology zoo examples that we annotated with business relationships
- **BGPStream examples**, annotated using the CAIDA AS relationships dataset
- [Gao and Rexford. SIGMETRICS 2000] Policy implements the Gao-Rexford conditions
- Properties: reachability (all-src), no-transit property

Machine details: 2.3 GHz Intel i7 processor, 16 GB RAM



### FatTree network with valley-free policy



- c = 0: route has 0 Aggr nodes
  c = 1: route has 1 Aggr nodes
  c = 2: route has 2 Aggr nodes
- c = 3: route has  $\geq$  3 Aggr nodes

#### Results for FatTrees with valley-free policy



- Abstract settings verify both properties without false alarms  $\bullet$
- abs\_mono verifies reachability for a FatTree with 36,980 routers in 40 mins
- lacksquare
- MonoSAT better for reachability but Z3 better for valley-free property  $\bullet$

Abstract settings are *uniformly* better than concrete settings (upto 52x speedup for MonoSAT)

### **Topology Zoo network: Example**



- Prefer Cust < Peer < Prov (Cust most preferred)</li>
- Don't export routes from Peer/Prov to another Peer/Prov (no-transit property)

42 nodes, 50 edges. Each node is an AS. Edges annotated with business relationships (customer/peer/provider)

[Gao and Rexford. SIGMETRICS 2000]

Policy implements Gao-Rexford conditions:







### **Results for Topology Zoo networks**



- Abstract settings verify both properties without false alarms
- Abstract settings are *uniformly* better than concrete settings (relative speedup of 3x)
- MonoSAT better for both properties



nout false alarms concrete settings (relative speedup of 3x)

#### **Results for BGPStream networks**



- Abstract settings verify no-transit property in all networks, and reachability in 6/10 networks ulletAbstract settings are better than concrete settings  $\bullet$
- - MonoSAT: speedup of 323x for reachability, and 120x for no-transit
  - Z3: benefit of NRC not as pronounced; no-abstraction setting sometimes better for no-transit  $\bullet$



#### **NRC Refinement**



- $\bullet$
- $\bullet$



4/10 false alarms are handled using a more precise abstraction that models local preference NRC abstraction hierarchy provides a tradeoff between performance and precision



### Comparison of ACORN with other tools



Reachability (single-src) on FatTree benchmarks with valley-free policy

- NV uses MTBDD-based simulation and SMT ShapeShifter uses BDD-based simulation with abstract interpretation
- Both NV and ShapeShifter *run out of memory* for networks with > 3000 nodes
- ACORN scales to  $\approx 37,000$  nodes
- SAT/SMT techniques seem more scalable than BDD-based methods (again!)

### **ACORN: Summary of experimental results**

- NRC can verify reachability for  $\approx 37,000$  routers within an hour • Far exceeds performance of existing control plane verifiers
- NRC improves scalability for both solvers, all benchmarks
  - Abstract settings uniformly better than concrete settings
- NRC could verify realistic policies

  - Some false alarms in WANs, refinements verified successfully
  - Future work: CEGAR-based refinement
- MonoSAT's graph theory solver useful for reachability
  - Z3 sometimes better for policy-based properties
  - This needs further investigation

• Common policies on data center networks, no false alarms with least precision



#### Lessons (re-)Learned

- Build the logic-based model first, leverage domain insights •
  - technology for verification
- Don't abstract too early
  - route redistribution) considered important by network practitioners
- on top of the logic-based model
  - Bonsai: Symmetry-based abstractions [Beckett et al. SIGCOMM 2018]
  - Shapeshifter: Abstract Interpretation [Beckett et al. POPL 2020]
  - Origami: Failure analysis [Giannarakis et al. CAV 2019]
  - NV: Programmable Platform [Giannarakis et al. PLDI 2020]
  - ACORN: Nondeterministic route selection [under submission]
  - Timepiece: Modular verification [under submission]

modeling the network control plane stable behavior enabled direct use of SMT

our SMT-based model is rich in detail, captures many features (e.g., local preferences,

Build logic-based abstractions, compositional methods, CEGAR, < your-favorite-method>





#### **Future Opportunities**

Many challenges still remain: scalability, failure analysis, ... •

- Quantitative/probabilistic properties
  - Some existing efforts (e.g., Probabilistic NetKat, ApproxFlow)
  - > Model counting techniques
  - Probabilistic verification
- Automated synthesis with verification  $\succ$  Leverage machine learning + deductive techniques



#### Collaborators







#### Ryan Beckett Microsoft Research

#### Ratul Mahajan Univ of Washington

#### Thank you!





#### Divya Raghunathan Princeton

#### Tim Thijm Princeton

Dave Walker Princeton

