RPKI Policy w/o Route Refresh

draft-ymbk-sidrops-rov-no-rr

NANOG 84

2022.02.14

Randy Bush, Keyur Patel, Philip Smith, & Mark Tinka

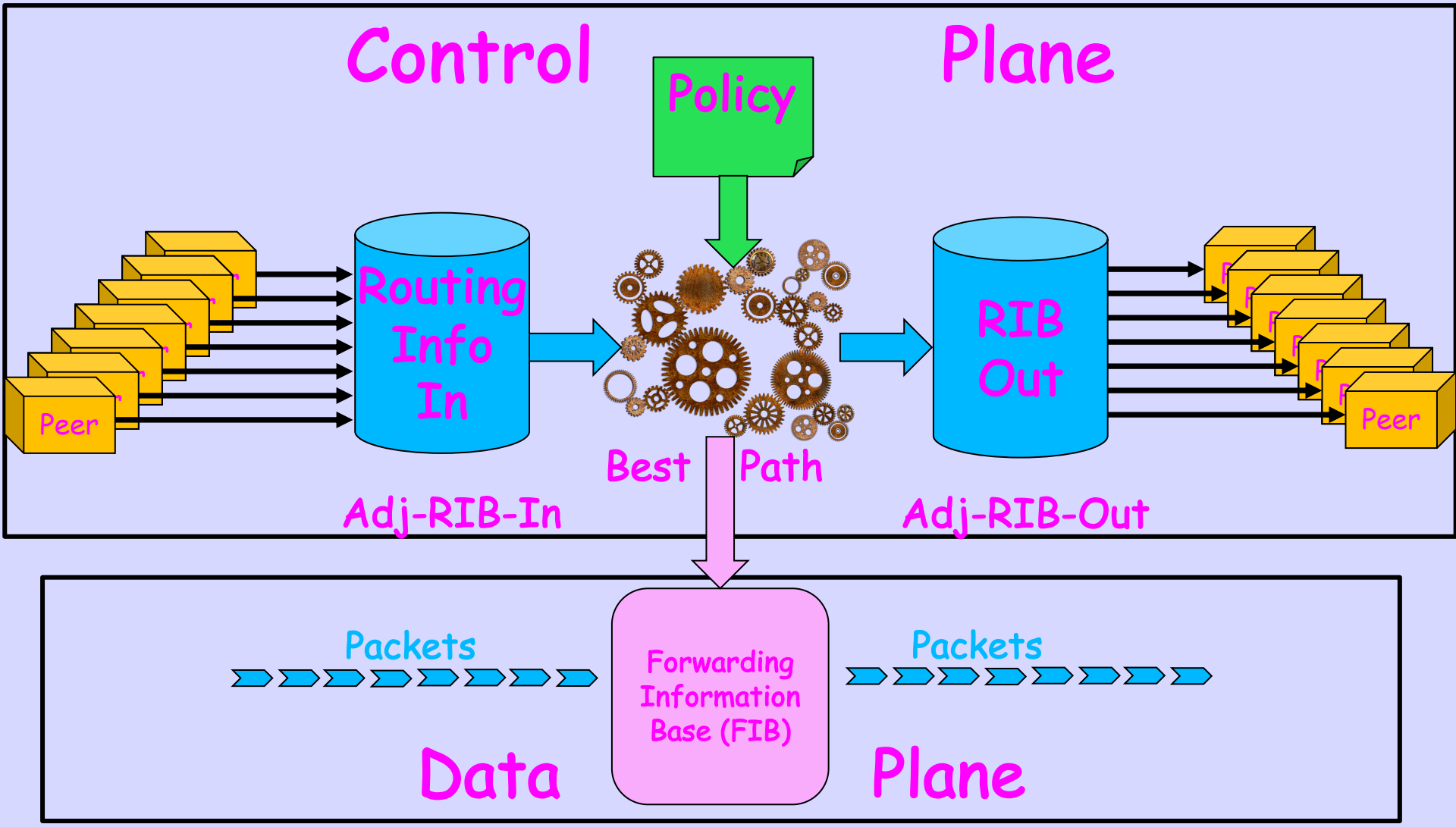
with John Heasley, Nick Hilliard, Ben Maddison, & John Scudder

Are you running BGP Route Origin Validation & dropping Invalids?

You may have some very annoyed BGP neighbors

They probably did not know this
They will now ☺

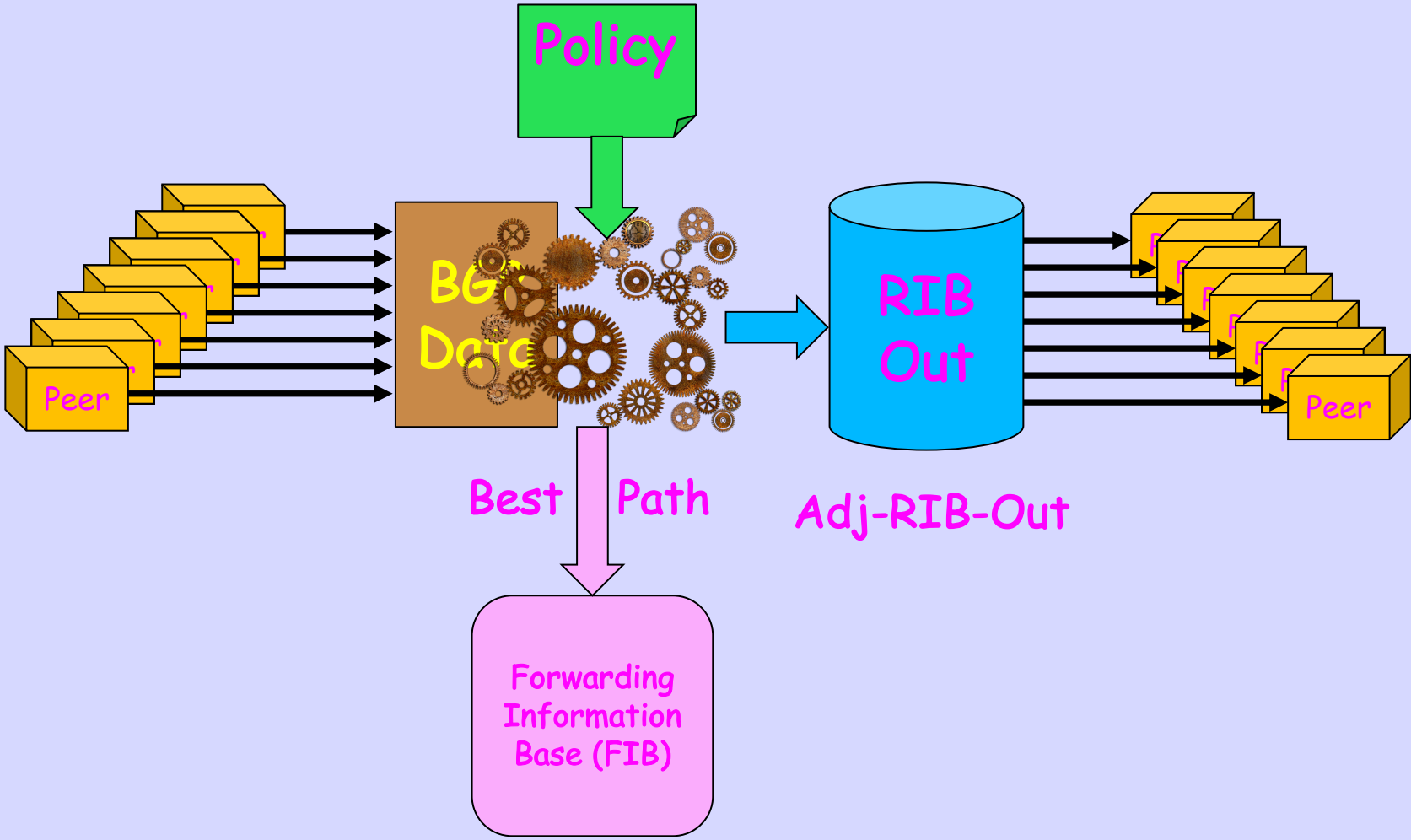
BGP Over-Simplified



No Adj-RIB-In

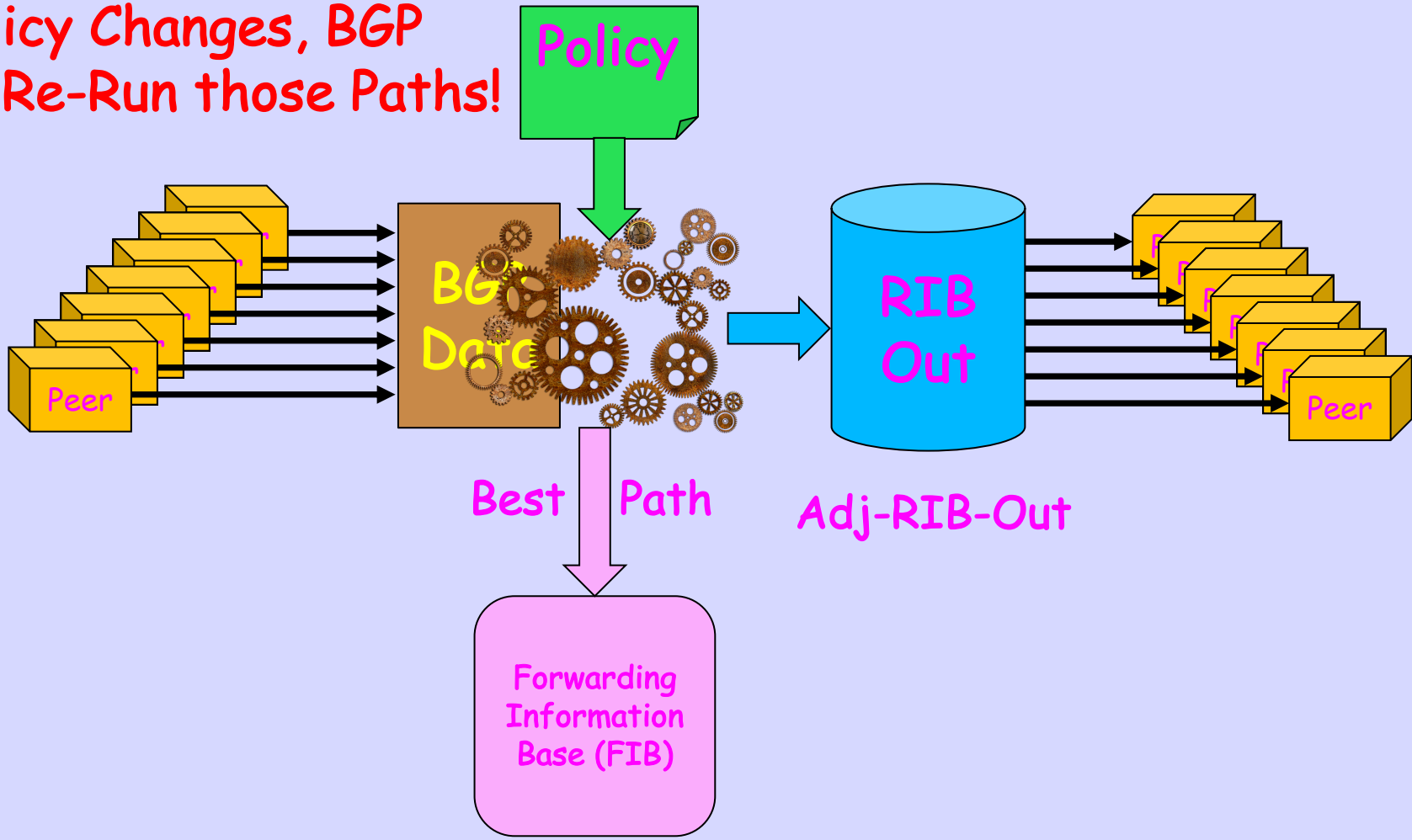
- It's the late-'80s, and RAM and CPU were horribly constrained
- Cisco's 004, Kirk Lougheed, develops data structures and algorithms which save memory by trading Adj-RIB-In/Out for internal data structures and traversals

BGP w/o Adj-RIB-In



BGP w/o Adj-RIB-In

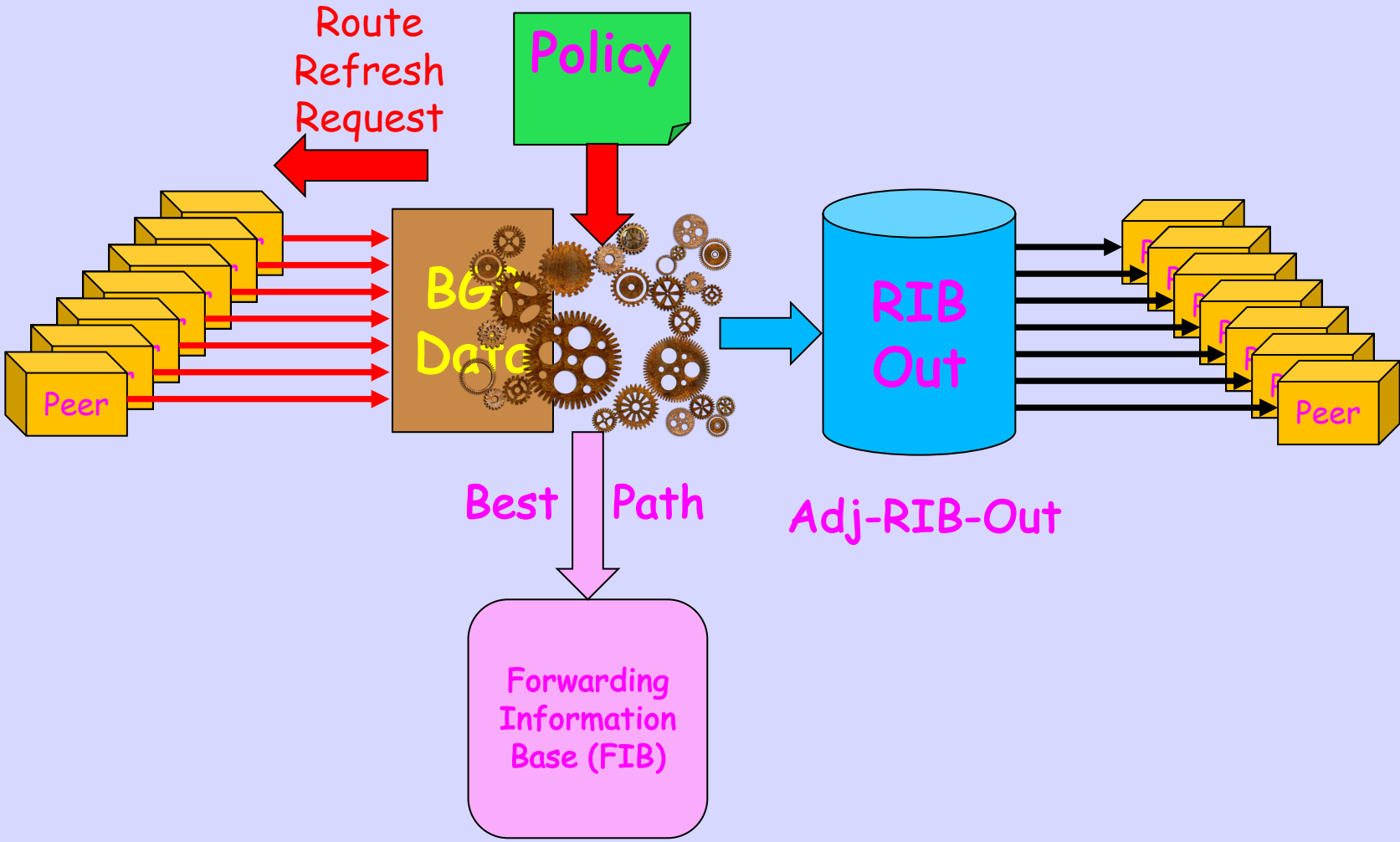
If Policy Changes, BGP Must Re-Run those Paths!



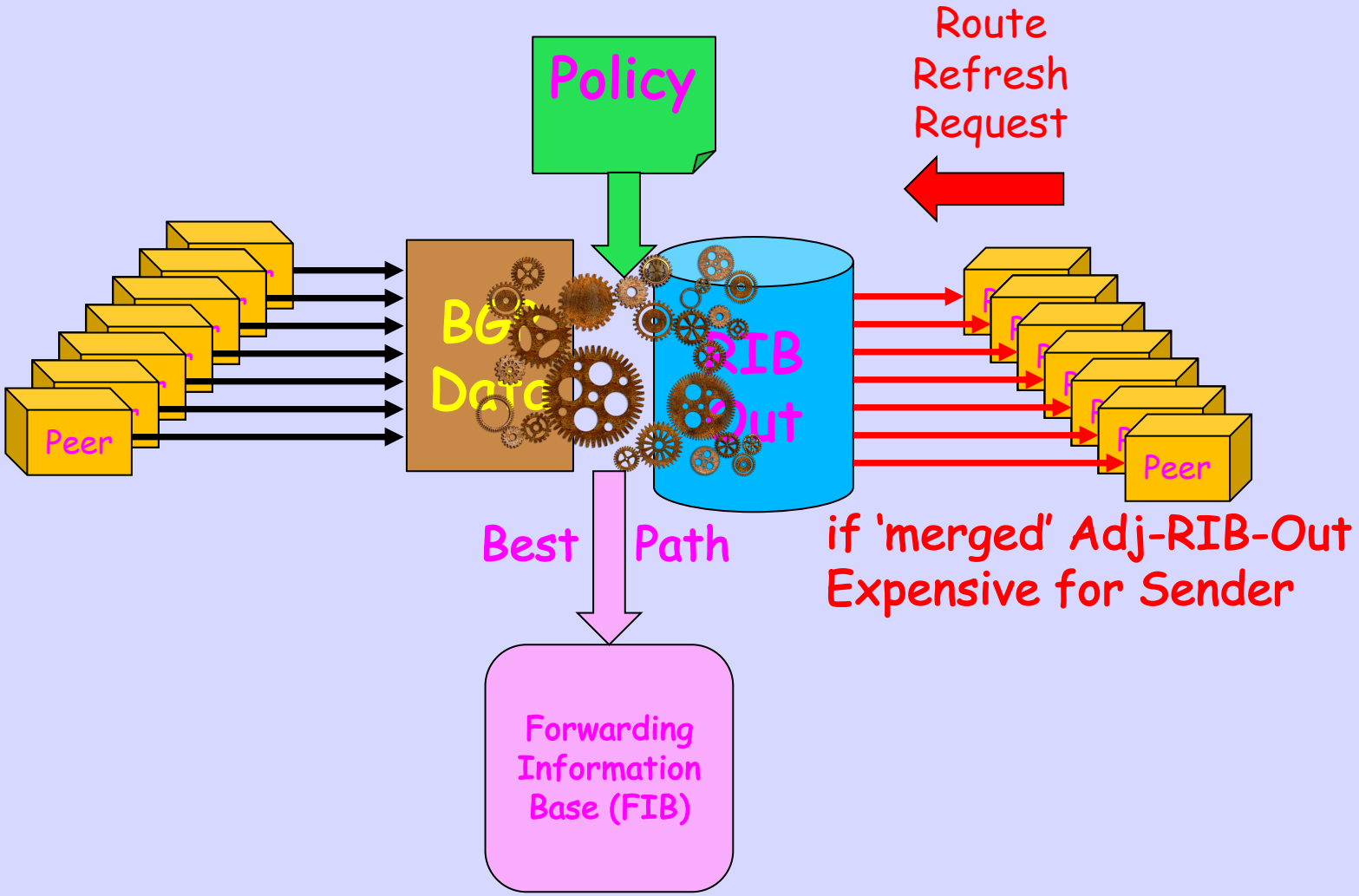
Route Refresh

- When Policy Changes, need to re-evaluate, but there are no Adj-RIB-In paths
- Enke Chen: *Route Refresh* to all Peers
- Peers resend all BGP paths to you, and you can run full policy
- And this substitutes for the lack of an Adj-RIB-In
- Policy changes rarely, so this is OK

Route Refresh

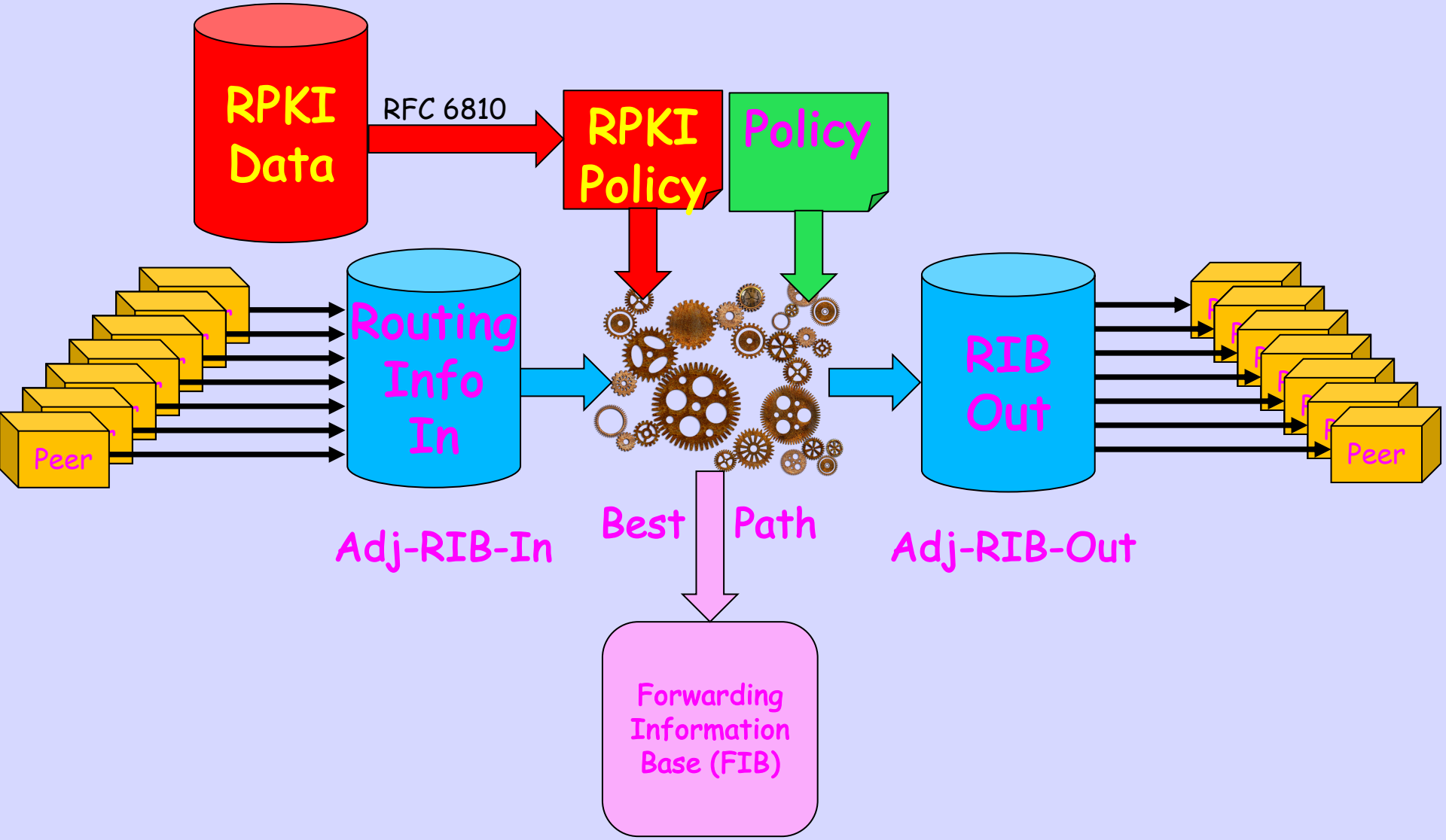


Route Refresh

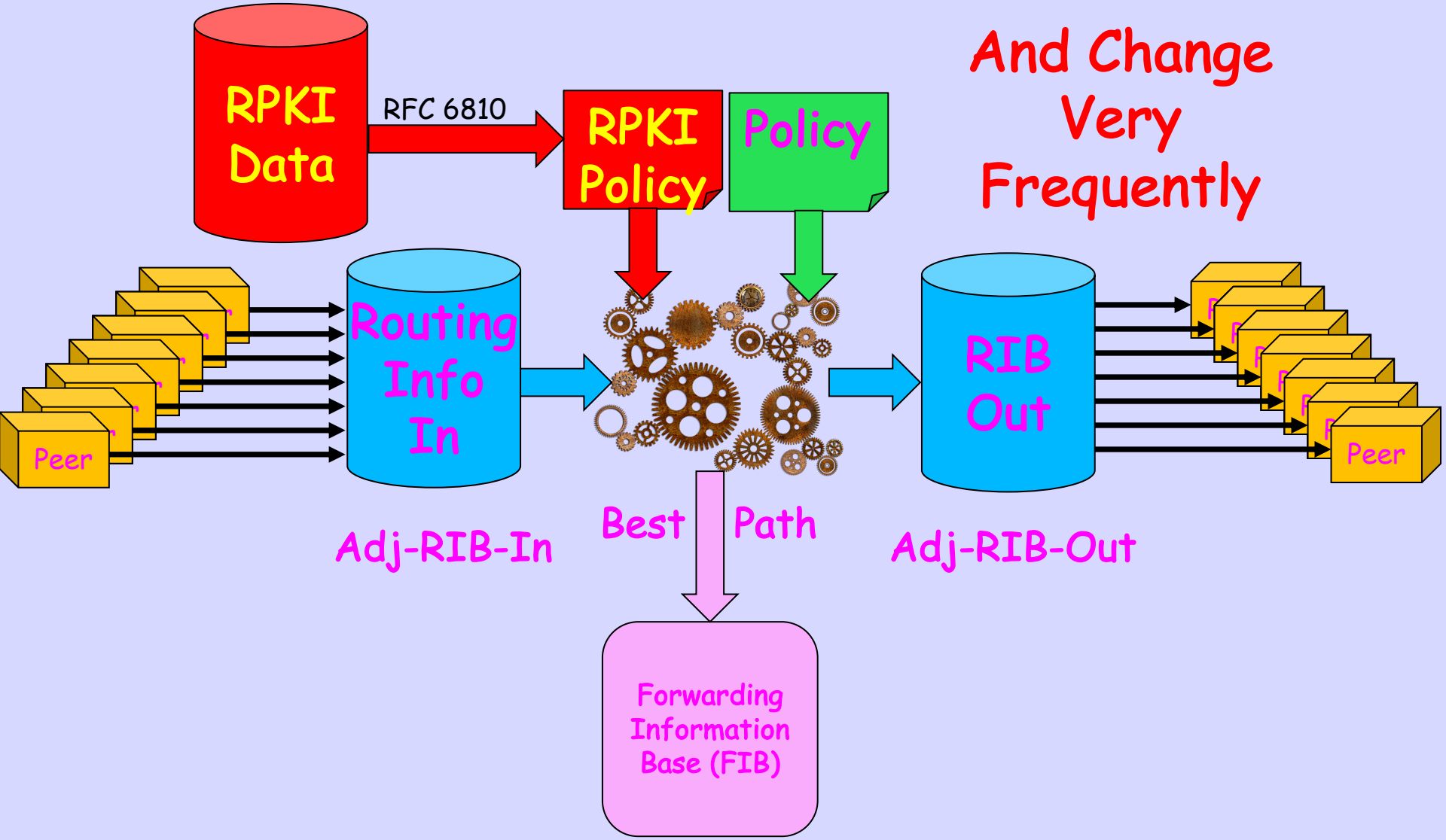


This works when Policy Changes Very Rarely

RPKI Data are New Policy

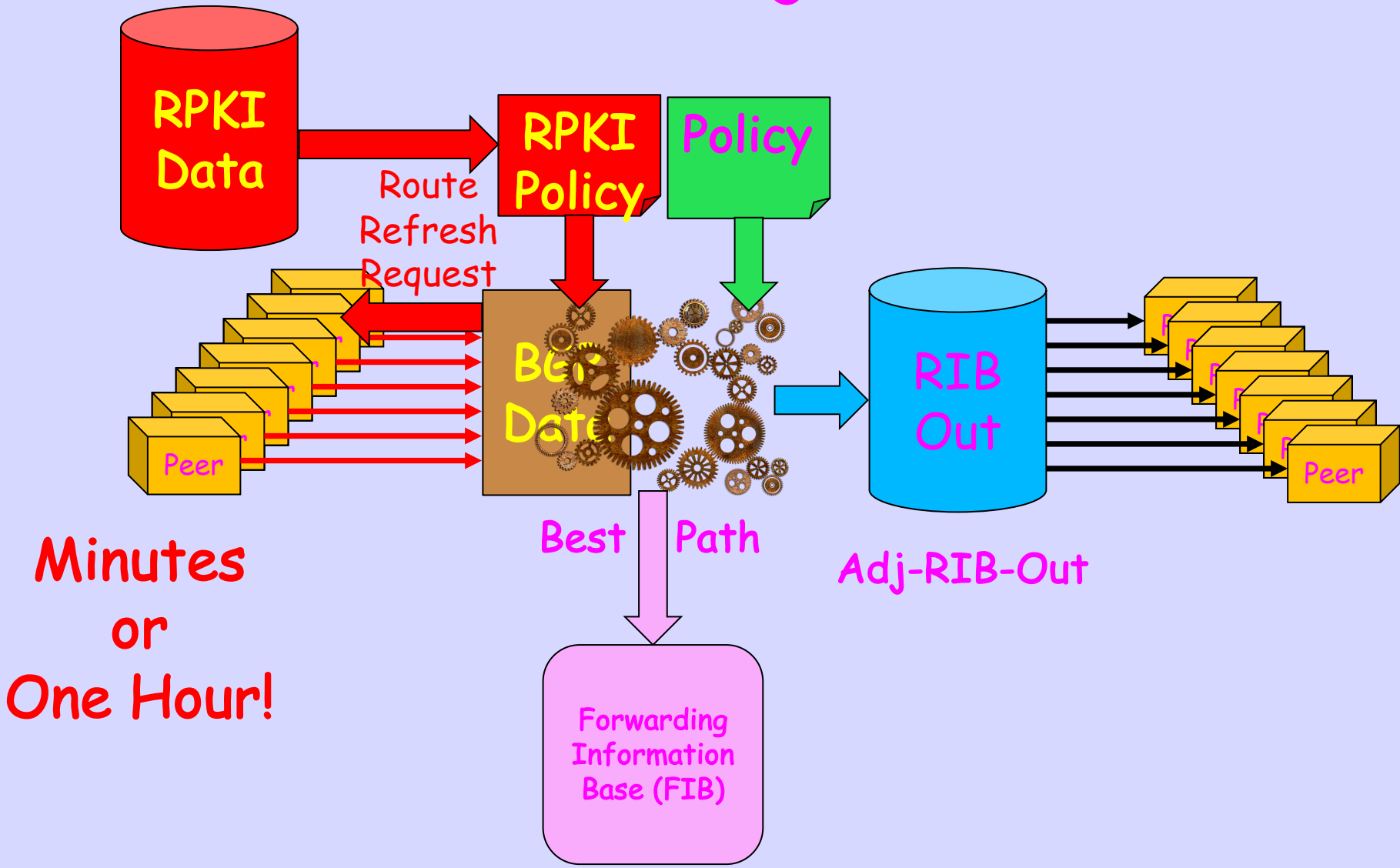


RPKI Data are New Policy



RPKI & Operator Policy Must be Done Before Best Path ReCalc

RPKI w/o Adj-RIB-In



Minutes or One Hour!

The Problem

- New RPKI Data require re-running RPKI & Operator Policy and BGP Best Path
- If there are no Adj-RIB-In data, prefixes dropped by previous RPKI policy (e.g. ROV Invalids) are no longer available
- So a Router issues a Route Refresh to all peers
- And this is NOT infrequent

Refresh Scales Poorly

- If you had Adj-RIB-In, a new ROA only affects a subtree, an operation with small scale

- Route Refresh gets a full table and the requestor must then re-evaluate the whole tree; a very expensive operation, with the side effect of Head Of Line Blocking for new incremental updates

And Worse (from jgs)

- An import policy change will often affect only a single peer and therefore require only a route-refresh solicitation toward that one peer
- A ROA change looks like a policy change toward all peers you're running ROV against, which is likely to be all your EBGP peers
- You have DDoSed yourself. Cool!

The Result

From: @att.com>
Sent: Friday, November 15, 2019 7:14 PM
To: Noah Maina; RUEEGG, DANIEL; peering; SEACOM INOC
Cc: jeffwei@juniper.net; bgtmo em
Subject: RE: Peer IP 80.: \\ AOTS – 270533338

Significant De-Peering!

Hello Noah/Seacom,

We have decided to shut these peering connections down.
AT&T in Frankfurt.

Consider an IXP with hundreds of members doing ROV and issuing Route Refresh to the Route Servers every few minutes.

Or the inverse, IXP Route Servers sending Route Refresh to hundreds of peers every time they get new RPKI data

Solution #1

- The obvious real solution is to keep a full Adj-RIB-In
- Some vendors do this by default, Yay!
- But that can be resource intensive on old hardware
- On modern hardware, there is no excuse not to keep Adj-RIB-In

Solution #2

- If no Adj-RIB-In, then when BGP drops an Invalid, keep the path, but mark it as dead, a minimal *Adj-RIB-Dropped*
- A lot smaller than full Adj-RIB-In
- Except if someone does another 7007 or UUNET 128/9, as those generate a 'jillion' dropped paths

Solution #3

- Do not run RPKI policy on any router which can not do #1 or #2
- Yes, we are telling you to turn off ROV dropping Invalids
- No, this does not make us happy

How to Test

- Different vendors have different ways to look at how often/much a router is issuing a Route Refresh
- The Route Refresh folk did not give us a MIB
- There are CLI and Yang queries on most devices to get RR counts

Why Didn't We Know This Was Happening?

We Do Not Measure Our Networks Well ☹

Questions