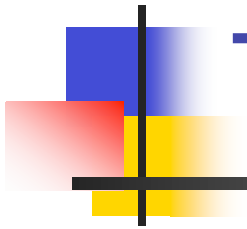


BGP Aggregation & The Deaggregation Report

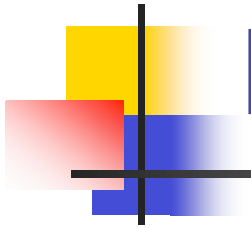
Philip Smith

JANOG 22
10th-11th July 2008

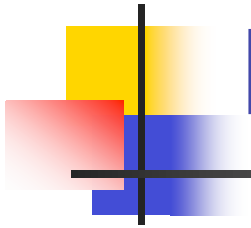
Route Aggregation Recommendations

- **LINXがIX接続メンバに対してaggregationポリシーの適応を試みた。**
  LINX attempted aggregation policy for members
  - **多くのメンバが支持したが失敗に**
    It failed even though most members voted for policy
- **2006年初めよりRIPE Routing WGのアイテムになる**
  RIPE Routing Working Group work item from early 2006
  - **最初のLINXのコンセプトをベースに**
    Based on early LINX concept
  - **執筆者：Philip Smith、Mike Hughes (LINX) 、Rob Evans (UKERNA)**
    Authored by Philip Smith, Mike Hughes (LINX) and Rob Evans (UKERNA)

Route Aggregation Recommendations



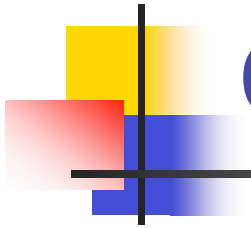
- **RIPE-399としてRIPEのドキュメントになる**
  RIPE Document — RIPE-399
  - http://www.ripe.net/ripe/docs/ripe-399.html
- **以下のディスカッション**
  Discusses:
  - 経路集約の歴史　History of aggregation
  - 経路細分化の原因　Causes of de-aggregation
  - グローバルな経路制御システムへのインパクト
    Impacts on global routing system
  - 現実的な解決策　Available Solutions
  - ISPへの薦め　Recommendations for ISPs



History:

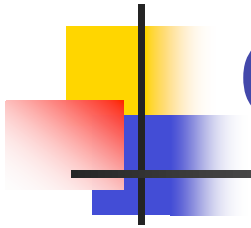
- クラスフルからクラスレスへの統合
  Classful to classless migration
  - 192/8を綺麗にしようプロジェクト
    Clean-up efforts in 192/8
- CIDRレポート CIDR Report
  - CIDRシステムとアグリゲーション適応の薦めとしてTony Batesが開始した
    Started by Tony Bates to encourage adoption of CIDR & aggregation
  - 90年代後半にかけてほとんど無視さてしまった
    Mostly ignored through late 90s
  - 現在は、Geoff Hustonによる広範囲にわたるBGP経路テーブルの解析として一部拡張されている
    Now part of extensive BGP table analysis by Geoff Huston
- RIRの概念やPAアドレス空間の導入
  Introduction of Regional Internet Registry system and PA address space

Deaggregation: Claimed causes (1):



- ルーティングシステムのセキュリティ
  Routing System Security
  - /24を広告するとlonger prefixを広告する"DOS"を防げる
    "Announcing /24s means that no one else can DOS the network"
- DOS攻撃や極悪非道な活動を軽減することができる
  Reduction of DOS attacks & miscreant activities
  - 実際使ってる空間のみを広告。大きい空間を広告するとごみトラフィックを吸い込んでしまうから
    "Announcing only address space in use as rest attracts 'noise'"
- 商用サービスでの理由　Commercial Reasons
  - 俺の勝手だろ？
    "Mind your own business"

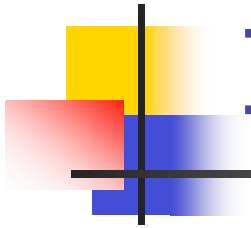
Deaggregation: Claimed causes (2):



- **iBGP localAS内部経路が外部に漏れるケース**
  Leakage of iBGP outside of local AS
  - **eBGPはiBGPではない – どのぐらいのISPはこれを知っている？**
    eBGP is NOT iBGP – how many ISPs know this?
- **マルチホーム接続におけるトラフィックエンジニアリング**
  Traffic Engineering for Multihoming
  - **/24をばらまいてなんとなくマルチホームした気になる**
    Spraying out /24s hoping it will work
  - **技術的な検討をちゃんとやるべき**
    Rather than do any **real engineering**
- **過去のレガシーな割り当て** Legacy Assignments
  - **全てのそれらpre-RIR割り当てには責任がある**
  - "All those pre-RIR assignments are to blame"
  - **実際には、RIRとレガシー割り当ての両方に当てはまる**
    In reality it is both RIR and legacy assignments

Impacts (1):

- ルータのメモリ　Router memory
  - ベンダがメモリ増加用件を過小評価するとルータのライフタイムを縮小してしまう
    Shortens router life time as vendors underestimate memory growth requirements
  - 減価償却のライフサイクルを短くする
    Depreciation life-cycle shortened
  - **ISP**、顧客のコスト増を引き起こす
    Increased costs for ISP and customers
- ルータの処理能力　Router processing power
  - プロセッサはベンダの最小評価**CPU**用件としても処理不足を引きおこす
    Processors are underpowered as vendors underestimate CPU requirement
  - 減価償却のライフサイクルを短くする
    Depreciation life-cycle shortened
  - **ISP**、顧客のコスト増を引き起こす
    Increased costs for ISP and customers



Impacts (2):

- ルーティングシステムの収束
  Routing System convergence
  - 経路表の増大 → 収束を遅くさせる
    Larger routing table → slowed convergence
  - コントロールプレーンのプロセッサの速度向上により改良される
    Can be improved by faster control plane processors — see earlier
- ネットワークのパフォーマンス & 安定性
  Network Performance & Stability
  - 収束が遅くなる → 故障からの回復が遅くなる
    Slowed convergence → slowed recovery from failure
  - 回復が遅くなる → ダウンタイムが長くなる
    Slowed recovery → longer downtime
  - ダウンタイムが長くなる → 顧客は喜ばない
    Longer downtime → unhappy customers

Solutions (1):

- CIDR Report
  - グローバルな経路集約に向けた努力
    Global aggregation efforts
  - **1994年以来レポーティングされている** Running since 1994
- Routing Table Report
  - **RIR毎の経路集約に向けた努力**
    Per RIR region aggregation efforts
  - **1999年以来レポーティングされている** Running since 1999
- Filtering recommendations
  - トレーニング、チュートリアル、**Cymru**プロジェクト
    Training, tutorials, Project Cymru,…
- “CIDR Police”

Solutions (2):

- BGP Features:
  - **NO_EXPORT コミュニティ** NO_EXPORT Community
  - **NOPEER コミュニティ** NOPEER Community
    - **1 ISPのみ実装している**
    - RFC3765 — but only one ISP has implemented it!!
  - **AS_PATHLIMIT アトリビュート** AS_PATHLIMIT attribute
    - IETFのIDRワーキンググループで現在も検討中
      Still working through IETF IDR Working Group
  - **プロバイダー固有のコミュニティ** Provider Specific Communities
    - **いくつかのISPは利用しているが、多くは利用していない**
      Some ISPs use them; most do not

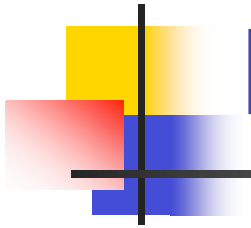
RIPE-399 Recommendations:

- 初期割り振りで受けた割り当て空間のみをシングルエントリーで経路広告する
  Announcement of initial allocation as a single entity
- もし連続した経路広告が可能なら、追加割り振りの際に経路集約をかける
  Subsequent allocations aggregated if they are contiguous and bit-wise aligned
- マルチホームネットワークに対しては集約経路の慎重な細分化が必要
  Prudent subdivision of aggregates for Multihoming
- 前述した**BGP**の機能を利用
  Use BGP enhancements already discussed
- （もちろん全て**がIPv6**にも適応されることが当然望まれる）
  (Oh, and all this applies to IPv6 too)

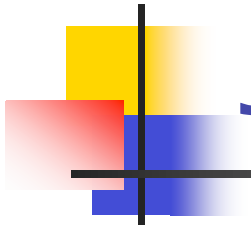
Looking at Deaggregation

- CIDR Report
  - www.cidr-report.org
  - InternetのCIDR化に従った経路集約の促進
    Encourages aggregation following CIDRisation of Internet
  - **グローバルBGPテーブルの状態を把握できる有効なツールやレポートが提供されている**
    Today: extensive suite of reports and tools covering state of BGP table
- Routing Report
  - **BGPテーブルサイズの状況をRIRごとにレポートしたもの**
    BGP table status on per RIR basis
  - **オリジナルのCIDRレポート及びより多くの範囲をカバー**
    Original CIDR Report and a whole lot more

Deaggregation Factor



- Routing Report
  - **originAS毎にBGPテーブルと集約経路**
  - One summary takes BGP table and aggregates prefixes by origin AS
    - **レポートでは、"MAX集約"と呼ばれている**
      Called "Max Aggregation" in report
  - **グローバル、RIR毎のレポート** Global and per RIR basis
    - http://thyme.apnic.net/current/
- New Deaggregation Factor:
  - **経路数 / 集約経路数 を計る**
    Measure of Routing Table size/Aggregated Size
  - **記録開始以来、グローバルの値は緩やかに着実に増加**
    Global value has been increasing slowly and steadily since "records began"



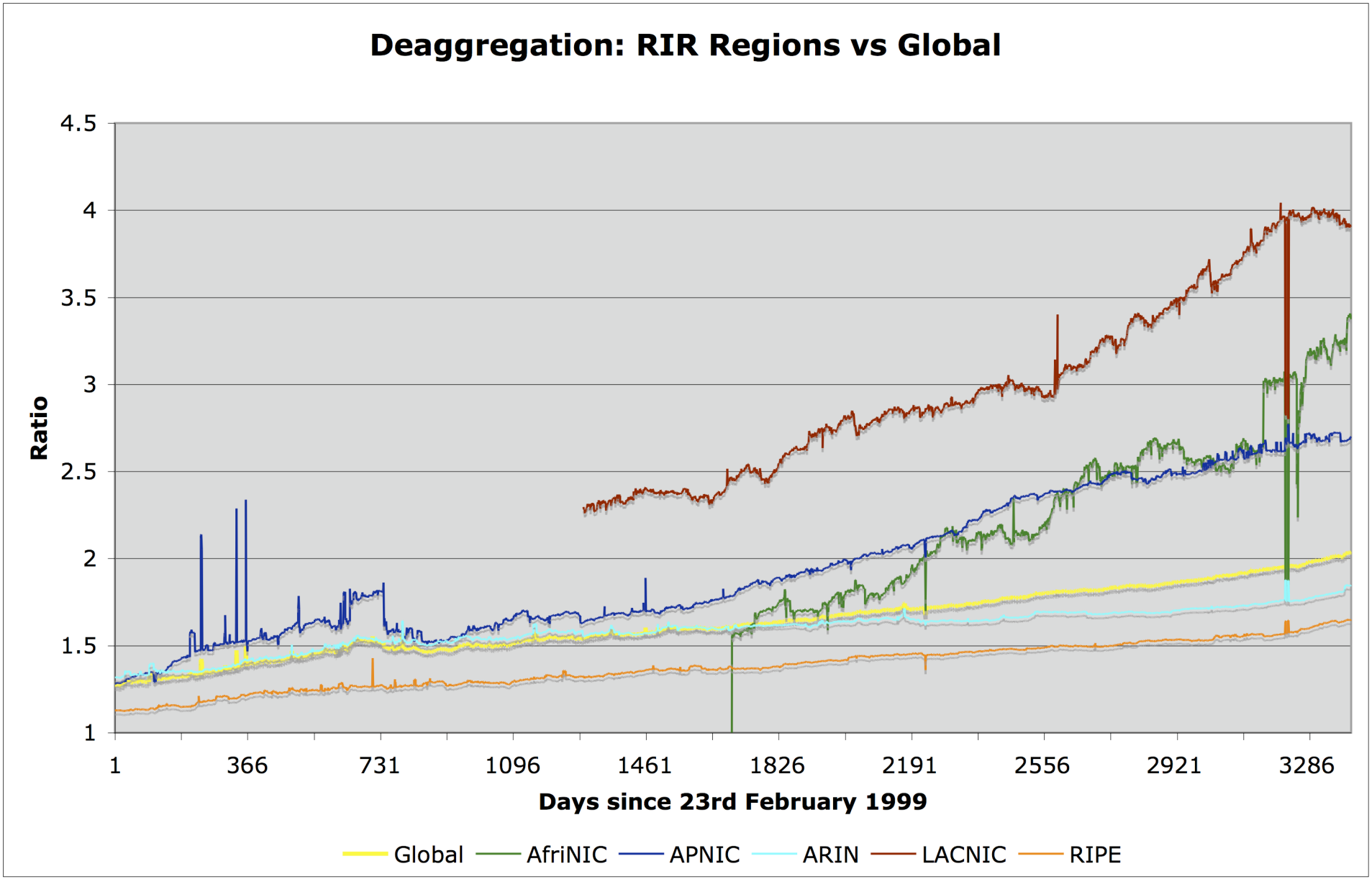
June 2008

Total Prefixes

- Global BGP Table
  - 261k prefixes
- Europe & Middle East
  - 56k prefixes
- North America
  - 120k prefixes
- Asia & Pacific
  - 60k prefixes
- Africa
  - 4k prefixes
- Latin America & Caribbean
  - 20k prefixes

Deaggregation Factor

- Global Average
  - 2.04
- Europe & Middle East
  - 1.65
- North America
  - 1.85
- Asia & Pacific
  - 2.69
- Africa
  - 3.36
- Latin America & Caribbean
  - 3.93


Deaggregation: RIR Regions vs Global
Ratio
4.5
4
3.5
3
2.5
2
1.5
1
1
366
731
1096
1461
1826
2191
2556
2921
3286
Days since 23rd February 1999
Global
AfriNIC
APNIC
ARIN
LACNIC
RIPE

Africa Aggregation Savings Summary

ASN	No of Nets	Poss Savings	Description
24863	475	445	LINKdotNET AS number
20858	397	394	EgyNet
6713	143	132	Itissalat Al-MAGHRIB
33783	135	123	EEPAD TISP TELECOM & INTERNET
2018	201	116	Tertiary Education Network
5536	121	105	Internet Egypt Network
29571	102	94	Ci Telecom Autonomous system
33776	99	91	Starcomms Nigeria Limited
24835	75	69	RAYA Telecom - Egypt
5713	155	62	Telkom SA Ltd
20484	63	60	Yalla Online Autonomous Syste
15475	63	59	Nile Online
15706	61	57	Sudatel Internet Exchange Aut
3741	273	49	The Internet Solution
29975	62	47	Vodacom
23889	68	45	MAURITIUS TELECOM
8094	42	39	PUKNET
16637	57	31	Johnnic e-Ventures
21152	32	31	AS for the uplinks of Soficom
12455	33	30	Jambonet Autonomous system

http://thyme.apnic.net/current/data-CIDRnet-AFRINIC

Asia & Pacific Aggregation Savings Summary

ASN	No of Nets	Poss Savings	Description
4755	1661	1485	Videsh Sanchar Nigam Ltd. Aut
17488	1188	1097	Hathway IP Over Cable Interne
9498	1079	1017	BHARTI BT INTERNET LTD.
9583	1157	739	Sify Limited
18101	686	652	Reliance Infocom Ltd Internet
4780	704	641	Digital United Inc.
9829	598	586	BSNL National Internet Backbo
4766	846	503	Korea Telecom (KIX)
4134	828	501	CHINANET-BACKBONE
17676	525	460	Softbank BB Corp.
7545	511	441	TPG Internet Pty Ltd
17974	456	439	PT TELEKOMUNIKASI INDONESIA
9443	468	394	Primus Telecommunications
4808	524	390	CNCGROUP IP network: China169
10091	341	330	SCV Broadband Access Provider
4668	333	326	LG-EDS Systems Inc.
4802	478	315	Wantree Development
23966	332	314	Dancom Pakistan (PVT) Limited
7552	296	292	Vietel Corporation
9304	300	268	Hutchison Telecom (HK)

http://thyme.apnic.net/current/data-CIDRnet-APNIC

North America Aggregation Savings Summary

ASN	No of Nets	Poss Savings	Description
6389	2670	2471	bellsouth.net, inc.
11492	1232	1220	Cable One
4323	1471	1094	Time Warner Telecom
18566	1045	1035	Covad Communications
1785	1080	976	AppliedTheory Corporation
22773	966	904	Cox Communications, Inc.
6478	956	779	AT&T Worldnet Services
19262	919	754	Verizon Global Networks
5668	694	661	CenturyTel Internet Holdings,
6517	700	653	Yipes Communications, Inc.
2386	1492	615	AT&T Data Communications Serv
3356	974	555	Level 3 Communications, LLC
855	598	545	Canadian Research Network
20115	1048	487	Charter Communications
19916	509	477	OLM LLC
6197	947	474	BellSouth Network Solutions,
7011	1015	461	Citizens Utilities
33588	447	421	Bresnan Communications, LLC.
7018	1395	419	AT&T WorldNet Services
8103	614	379	Florida Department of Managem

http://thyme.apnic.net/current/data-CIDRnet-ARIN

Latin America Aggregation Savings Summary

ASN	No of Nets	Poss Savings	Description
8151	1273	1046	UniNet S.A. de C.V.
11830	604	595	Instituto Costarricense de El
22047	565	551	VTR PUNTO NET S.A.
16814	426	416	NSS, S.A.
7303	469	404	Telecom Argentina Stet-France
14117	375	366	Telefonica del Sur S.A.
6471	411	363	ENTEL CHILE S.A.
11172	410	340	Servicios Alestra S.A de C.V
10620	404	339	TVCABLE BOGOTA
10481	310	301	Prima S.A.
28573	303	274	NET Servicos de Comunicao S.A
20299	335	237	NEWCOM AMERICAS
14259	296	235	GTD Internet S.A.
7738	252	226	Telecomunicacoes da Bahia S.A
14522	194	186	SatNet S.A.
19169	205	184	Telconet
23216	243	183	RAMtelecom Telecomunicaciones
8163	187	174	METROTEL REDES S.A.
21826	205	164	INTERCABLE
6458	173	157	GUATEL

http://thyme.apnic.net/current/data-CIDRnet-LACNIC

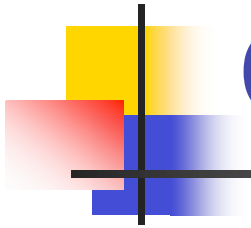
EU & Middle East Aggregation Savings Summary

ASN	No of Nets	Poss Savings	Description
8452	347	336	TEDATA
8866	319	298	Bulgarian Telecommunication C
5462	296	269	Telewest Broadband
9155	265	253	QualityNet AS number
8551	287	249	Bezeq International
12479	229	223	Uni2 Autonomous System
9121	249	222	TTnet Autonomous System
29357	216	212	WATANIYA TELECOM
3352	246	204	Ibernet, Internet Access Netw
35141	206	200	Megalan Autonomous system of
3215	286	197	France Telecom Transpac
9198	204	194	Kazakhtelecom Data Network Ad
3269	241	169	TELECOM ITALIA
6830	187	145	UPC Distribution Services
9051	160	138	INCONET Autonomous System
3300	231	132	AUCS Communications Services
8877	137	130	BOL.BG Autonomous System
29314	148	129	Telewizja Kablowa Dami Sp. z
5486	140	123	Euronet Digital Communication
1267	156	119	Infostrada S.p.A.

http://thyme.apnic.net/current/data-CIDRnet-RIPE

Observations

- **RIR地域の運用における慣例の範囲**
  Range of operational "practices" between RIR regions
  - **インターネット接続経験が浅い地域（Newer Internet）が急速に発展**
    "Newer" Internet is growing rapidly
    - **deaggregationが示しているように**
      As is the deaggregation there
- **RIPE-399は、推奨でしかない**
  RIPE-399 is only a recommendation
  - **できれば、各々の割り振り毎に全RIRがポインターに含めてくれることを望む**
    Hopefully all the RIRs will include pointers with each address allocation
  - **できれば、もっとより多くのISPが注目してくれることを望む**
    Hopefully more ISPs will pay attention to it
  - **トレイニングはここにあり ― 多くのISPは無視することを選択している**
  - Training is there — most ISPs choose to ignore it



Conclusion

- **RIPE-399 を是非オペレーションバイブルに！**
  Make RIPE-399 your BGP good practice document