Public Internet Routing Registries (IRR) Evolution

Akmal Khan,Hyunchul Kim,Ted "Taekyoung" Kwon & Yanghee Choi
Seoul National University (Korea)
{raoakhan,hkim,tk}@mmlab.snu.ac.kr,ychoi@snu.ac.kr

ABSTRACT

Internet Routing Registries(IRR) have been around for quite some time now[1] with the sole purpose of providing the place for service providers to store their administrative, routing policy information which can be used in case of BGP malicious/misconfiguration events. Are there any useful service providers policy data stored in IRR? What current limited research has able to answer is that "Quality" of IRR databases is not known". By "Quality" we mean validity of Internet Number Resources e.g. IPv4, IPv6, AS Number registration, routing policy registration, etc by different network service providers in IRR. We have tried to answer this question by looking into the public IRR datasets of approximately last 4 years [2006-2010].We have found out that current IRR datasets has a lot to offer than its known/practiced i.e. IRR has approximately 50k full peering available. We are investigating how many peering are in harmony with what BGP is announcing and also which is currently published in well known topology datasets like UCLA IRL[25].As we believe that if accurate peering can be extracted from IRR than they can provide number of new links which are missing in Internet Topology datasets. It can also reduce the usage of active measurements which in itself is burden on the network. We are also designing BGP Security framework based on IRR which will more accurately perform origin AS authentication as well as inferring the complete policy(what is stored in IRR) of AS.

Categories and Subject Descriptors

C.2.0 [**Computer-Communication Networks**]: General *Security and Protection*; C.2.2 [**Computer-Communication Networks**]: Network Protocols—*Routing Protocols*; C.2.3 [**Computer-Communication Networks**]: Network Operations --*Network Monitoring*

Keywords

Inter Domain Routing, BGP; Internet Routing Registries

1. INTRODUCTION

The Border Gateway protocol (BGP) [3] is the glue which helps Internet to provide its numerous services and unfortunately also the most fragile and vulnerable component of global routing system [11, 22]. The Border Gateway Protocol (BGP) is used to exchange destination reach ability information (routes) between different autonomous systems (AS) on the Internet. These routes consist of blocks of IP addresses, often referred to as prefixes, which are allocated to ISPs or end sites by RIRs (Regional Internet Registries). Beyond allocating addresses, RIRs play no actual role in the operational aspects of Internet routing [23]. Major issues lingering over BGP are related to BGP Route Hijacking whether malicious or unintentional. There are number of solutions over the years to target the BGP security issues like prefix hijacking, AS path spoofing but none of them have been deployed due to their limitations and current business practices [10, 11].

Researchers have considered the number of data sources to be used for their BGP Security solutions for prefix hijacking, AS path spoofing, etc. One of major data source is BGP traces collected by number of BGP collectors [19, 20].BGP collected traces have been used in number of research projects but on the other hand another viable data source i.e. Regional Internet Registry (RIR) allocation and policy data sets which are published as part of their WHOIS database and formally called as Internet Routing Registries (IRR).IRR routing information has not been used to the same extent (as BGP) due to known/assumed limitations of inconsistent data 1) static/voluntary nature of IRR) 2) stale, overlapping 3) incomplete information [4, 9].Current research has helped raised the issues surrounding IRR data but there are limited open source tools available which can help extract useful data from IRR [9,10,11]. There currently exists no strict linkage between RIRs allocation and IRR with the exception of some Réseaux IP Européens (RIPE NCC) mechanisms [6], Nor does there exists any linkage between RIRs and the actual routing system itself. Network Service Providers mainly employs IRR for routing policy generation and filtering of customers and peers [22].

Established research on IRR can be classified into projects checking the consistency of IRR data[4,6,9],using IRR data in generating the most accurate topology[7,8],and proposed scheme for using IRR data as a useful service to solve some of the BGP security issues [10,11].IRR consistency based research has reported that RIPE is the most consistent dataset and APNIC is also not far behind.[9] Unfortunately only the dataset from [4] has been published and available to researchers for analysis. There is no recent work which can validate the results published in 2004 and 2006.Topology based research results states the usefulness of IRR data in providing unique links but it's quite hard to prove the authenticity of those links due to possible stale IRR datasets. Some of the issues of IRR dataset which past research has found out are related to Number of objects registered in more than one registry i.e. overlapping information,

CFI'10 June 16-18, 2010, Seoul, Korea.

timestamps stored with objects shows quite old dates indicating the staleness of routing information ,etc. Current proposed algorithms have also failed to include the (1) Multi homing aspects 2) How community attributes are used 3) Using only limited RPSL constructs e.g. Autnum, Inetnum, etc. There is no tool available at this time which could able to infer the complete policy of an Autonomous System(AS).By complete policy we mean ,AS registered policy in the IRR which can also be confirmed from its neighbors ASes records in IRR.

We have analyzed the evolution of IRR dataset of the last 4 years i.e. From March 2006-March 2010.We are interested in knowing 1)How IRR have evolved in this time period? 2) How much data have been fed into IRR? 3) What's the accuracy of this datasets (operational uses) 4) How much complete information is stored in IRR databases? i.e. in comparison with different RPSL objects.

Our proposed contributions are as follows:

1. We have evaluated the IRR evolution in past 4 years to know the increase/decrease in the way different RPSL objects are entered in IRR. We reaffirm the belief in the research community about the operational practice of IRR's that Route objects are the mostly entered RPSL constructs by different AS's.
2. We are working on to answer the question: How much useful an up-to-date registry can be in solving the problems of Internet Routing? If RIPE NCC registry is good than Why it's not comparable to BGP traces. How can than we say that it is good? Why do we find topology differences between BGP and IRR as reported in [7]?
3. We are also working on to follow a Community-centric view of analyzing AS policies rather than AS-centric view which has limited view when considering routing policies of more than one hop neighbors. We process and clean the information in order to minimize the effect of inaccurate information.
4. We are working on BGP Security framework which includes more accurate method of validating the origin AS. We rely on currently available information, mainly the public Internet registries like RIPE, RADb.

In this paper we are only presenting results related to our 1st contribution i.e. IRR evolution. The rest of this paper is structured as follows. In section 2 we present some background and related work. In section 3, we present IRR evolution and some analysis results and; in section 4 we present our conclusions.

2. BACKGROUND & RELATED WORK

Administrative procedures are necessary to ensure the uniqueness of the IP addresses and Autonomous System numbers. The Internet Assigned Numbers Authority (IANA) [12] is responsible for global coordination of key elements for running the Internet smoothly. There are 5 Regional Internet Registries (RIR) namely RIPE NCC, APNIC, ARIN, LACNIC, AFRINIC [12]. There are two types of IRRs' namely public and private IRR's. One of the popular public registry; Internet Routing Registry (IRR or RADb) [14] has been setup since 1995 and it currently has 32 registries. RADb [15] and RIPE NCC [16] are public Internet Registries which publishes the IRR database daily. Different IRRs can manage their databases independently and also exchange registry objects database on regular basis. Some of the goals of IRRs are to provide network operators to share routing policy information e.g. when establishing peerings, Network troubleshooting, achieving stability and consistency of routing through publicly announced routing policies, etc. It could also be helpful in case of loss of hardware/administrators which results in less downtime. But unfortunately little has been achieved so far [1, 2].Over the years RIPE NCC [13] has played a key role in taking the idea of IRR moving forward with the policy specification published as RIPE-81 and later RIPE-181.Latest IETF proposed standard based on RIPE-181 published in 1999 as Routing Policy Specification Language (RPSL)[RFC 2622][1] which is an object based language consisting of 13 classes. RPSL Classes [1] can be divided into two main groups of policy and administrative objects. Some of the advantages of using RPSL as routing policy language are that it is extensible, not vendor specific and it has global view rather than router specific view of policies. RPSLng [RFC 4012][2] added support for IPv6 and multicast while Route policy System Security [RPSS][RFC-2725] added security extension to RPSL [5].RPSL considers the IRR system as a whole but in reality it consists of number of registries maintaining their policy records publicly or privately. This has led to number of issues including not able to achieve the full implementation scale of IRR's idea.

RIPE NCC has published RIPE database which is currently in version 3.9.RIPE routing registry is subset of the RIPE database and holds routing information in extended version of RPSL.RIPE database currently supports 21 objects which are all those that are defined in RPSL Specification[1] and in addition as-block,domain,keycert,limerick,irt,organization,route6,poem, poetic-form classes. Every class supports number of attributes ranging from to uniquely identifying the objects known as class "key". A mandatory attribute has to be defined for all objects of the class; optional attributes can be skipped. Attributes can also be single or multiple valued. Detail information about RPSL constructs, sample usage can be accessed from [1, 2].The accurate, up-to-date maintenance of the RPSL database can help contribute toward such goals as router configurations that protect against accidental (or malicious) distribution of inaccurate routing information, verification of Internet's routing, and aggregation boundaries beyond a single AS.

Routing Registry Consistency Check (RRCC) [6] Project by RIPE is one of the first to focus on the goals of making RIPE IRR useful to network engineers. RRCC provides a service where network engineers can check the accuracy of Routing Registry against the data with the actual routing announcements from their networks. Comparisons like AS's route objects registered and actually announced, non-registered peering detection which is done using the data provided by another RIPE project Routing Information System [RIS] [20]. Siganos et al.[9] developed a tool, called Nemecis that checks the correctness of IRR data and their consistency with respect to BGP routing table information. They argued that 28% of ASes have both correct and consistent policies and that RIPE is by far the most accurate registry. TERRAIN [11] is the most recent work which incorporates BGP traces and IRR data to propose BGP Security solutions for prefix hijacking, Origin Autonomous System (AS) Authentication.

Table 1 : Most populated Public Internet Routing Registry of March-2006 & March 2010

	APNIC		RIPE NCC		ARIN		Bell		Level3		RADb		NTTCOM	
	03/2006	03/2010	03/2006	03/2010	03/2006	03/2010	03/2006	03/2010	03/2006	03/2010	03/2006	03/2010	03/2006	03/2010
AUT-NUM	495	5,420	11,468	19.905k	324	1014	74	92	75	245	608	3,253	498	545
ROUTE	16	50,779	55,793	119.359k	194	10959	26049	29463	498	76,325	121	2,98,378	26229	75,156
INETNUM	0	0	1580686	2911.238k	0	461	0	0	0	0	0	0	0	0
INET6NUM	0	0	10,986	35.303k	0	1	0	0	0	0	0	0	0	22
AS-SET	2	198	5172	9.422k	14	298	49	50	2	188	26	1,666	340	423
FILTER-SET	0	0	63	0.096k	0	0	0	0	0	0	0	12	0	0
PEERING-SET	0	3	143	0.171k	0	0	0	0	0	0	0	13	0	0
ROUTE-SET	4	43	578	0.983k	6	208	3	3	10	1658	35	1654	228	234
RTR-SET	0	0	11	0.014k	0	0	0	0	0	0	0	9	0	0
KEY-CERT	7	8	3299	6.071k	0	0	0	0	0	0	41	248	21	653
ROUTE6	0	61	219	0	0	10	0	92	0	10	0	637	54	83
INET-RTR	0	0	100	0.114k	0	0	0	0	0	2	0	95	1	1
Total	**514**	**>56K**	**>1.6M**	**>3M**	**538**	**>12K**	**>26K**	**>29K**	**585**	**>78<**	**831**	**>300K**	**>27K**	**>77K**

3. Internet Routing Registry Evolution

The most recent work on IRR Analysis is by Battista et al.[4]. IRR Analysis Service provides an online service for checking the consistency of IRR. It does that by providing General statistics on the IRR (number of objects defined in each registry, amount of overlapping information between registries, etc.), set of pairs of ASes corresponding to peering relationships extracted from the IRR. Each pair is labeled with information about the context where it has been found, like the type of policy and the registry. We would like to emphasis the point that there is no other service which is providing this kind of useful service for the research community. As we have already mentioned that there are two well known public registries i.e. RADb and RIPE NCC. Both services publish the daily snapshots of their database so there is no way for researchers to look at the historical evolution of IRR. Some of the issues we have found out are that IRR analysis service scripts broke down since June 2008 and not reporting accurate IRR analysis. For our IRR Evolution analysis we have collected their datasets which they are continuously publishing since 2006.We have also collected the datasets from RIPE NCC. One of our observations from the datasets is that RADb had 68 registries published in 2006 which has reduced to 32 in 2010, which also includes some new registries. Where those published registries in 2006 has gone? One of the possible reasons is that some of the registries have stopped mirroring their information in RADb, instead of maintaining separate registry; they have started registering their objects in RADb registry etc.

Table1; shows the snapshot of most populated registries from RADb and RIPE NCC of two months i.e. March 2006/March 2010.RIPE NCC has >3M objects and RADb has >300K objects . We are not reporting here the RPSL objects like mntner, person, role as they are mostly required to authenticate while registering data into IRR. Some of our findings are 1)RIPE NCC is the mostly populated registry as well as usage of different RPSL constructs while other registries still use only partial set of RPSL constructs 2).autnum and route objects are mostly published objects in IRR so any scheme which relies on using other objects has limited success in terms of using published IRR datasets.3) Verio has changed its name to NTTCOM and contains around 77K objects which clearly shows that policies set by service providers can push customers to enter data into IRR. Figure 1 and Figure 2 shows the evolution of different RPSL objects in our observed period of March 2006- June 2008.

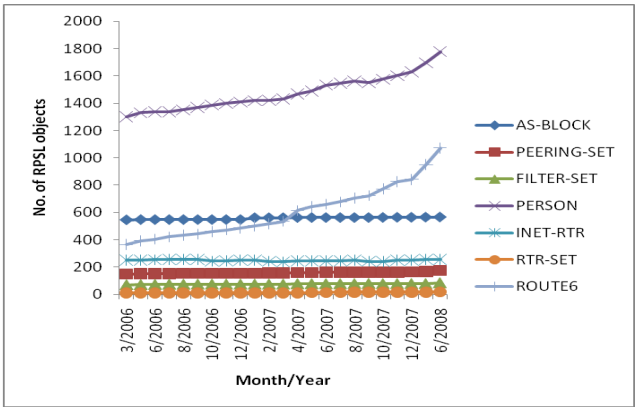

Figure1: RPSL Objects Evolution (Minimal Entry)

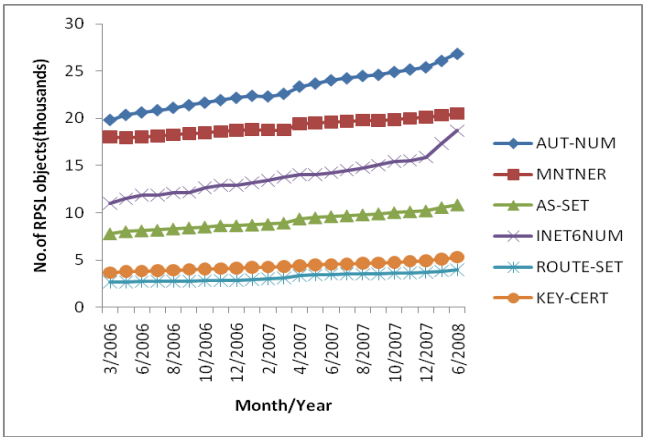

Figure 2: RPSL Objects Evolution(less actively entered)

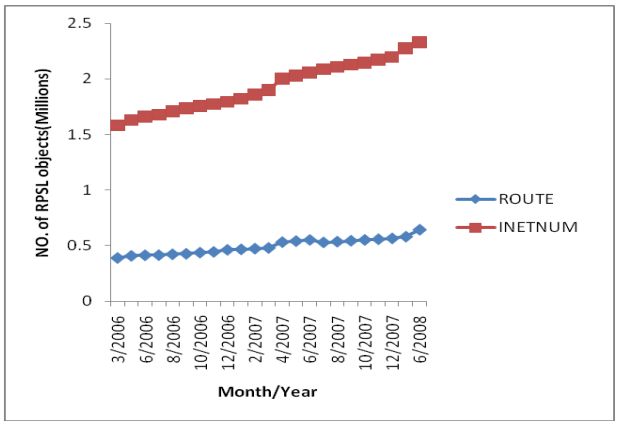

Figure 1: RPSL Objects Evolution (Most Actively Entered)

Figure1-3; shows the count of number of total RPSL objects existence in IRR.As it can be seen that operators mostly registered object registered in IRR's are Inetnum, Route and AS-set. Rest of the objects has not been used or some of the objects are shared between different AS's. Most of the policy data which is pushed if at all into IRR's are to satisfy the requirements of some providers by Customer ASes. Network Service providers can automatically generate router configurations from customers IRR records.

IRR Analysis Service [4] also provides the peering extracted from IRR dataset. They have also shared with us their service scripts. We are evaluating these scripts and will make comparison in our later publications. We are working on to generate full AS policy which an AS has registered in IRR.IRR analysis service extracts from the IRR the peering relationships between ASes by analyzing the body of RPSL objects. It than classifies the computed candidate peerings in order to understand the extent they contribute to fully specifies a peering. There are number of different peering types defined in dataset but we have only included the 5 peering types which are mostly found in IRR.

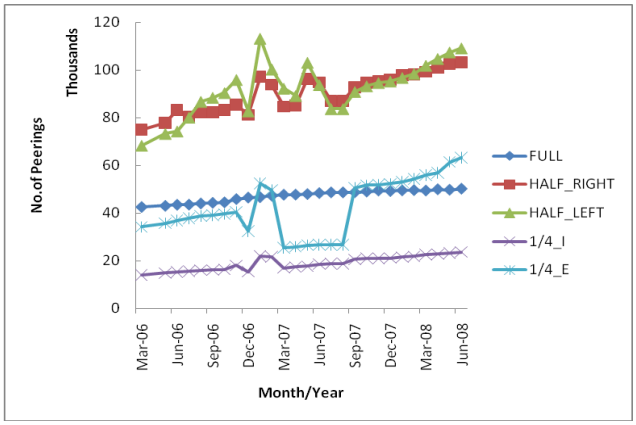

Figure 4: Major Peering Types existence in IRR

Figure 4; shows the increase in different type of peerings existence in IRR datasets from March 2006 to May 2008.We have not able to get data after May 2008 due to broken scripts from IRR analysis services. We have taken the terminology defined by IRR analysis service [4] to describe different type of peerings. For example there are two AS's A and B. there can be 4 records to supports their peering relationship i.e. A export B, A import B,B export A and B import A. By "FULL" peering type it means that Both AS's RPSL records confirms the peerings."HALF_RIGHT" means that from RPSL records what we confirm is (A import B and B export A)."HALF_LEFT" means we have able to extract (A export B and B import A) ."1/4_I" means that we have only able to confirm (A import B or B import A)."1/4_E has the same meaning as "1/4_I" except it specifies export rules.

We are working on our BGP Security Scheme that uses IRR data along with BGP traces history to help mitigate the prefix hijacking and AS Path spoofing problems. Despite its limitations IRR data can be used to solve some of the operational problems of BGP but there is an urgent need to define proper IRR maintenance policies by RIR.

4. CONCLUSION & FUTURE WORK

We have tried to raise the issues related to the importance of IRR by presenting how the size, accuracy of information stored in it has evolved over the years. There is strand of research available on IRR which has raised the issues but have failed to provide the community with useful tools.[IRR analysis Service [4] is an exception].We are interested in extending the work on IRR by publishing/extending available tools/methodologies as it can solve some of the issues like BGP prefix hijacking/misconfiguration. Research on IRR data can be of great help in term of solving the policy related issues by different Network Service providers, can be used as a "ground truth" as there are very limited information provided by Network Service Providers about their network policies, dataset for Network topology research as it has used in very limited way due to difficulty of extracting/validating links information in IRR's.

5. ACKNOWLEDGEMENT

This work was supported by NAP of Korea Research Council of Fundamental Science & Technology. The ICT at Seoul National University provides research facilities for this study

6. REFERENCES

1) C. Alaettinoglu et. Al Routing Policy Specification Language (RPSL). IETF RFC 2622, 1999.
2) L. Blunk, J. Damas, F. Parent, and A. Robachevsky.Routing Policy Specification Language next generation (RPSLng). IETF RFC 4012, 2005.
3) Y. Rekhter, T. Li, and S. Hares. A Border Gateway Protocol 4 (BGP-4). IETF RFC 4271, 2006.
4) G.Di.Battista , T.Refice , M.Rimondini, How to extract BGP peering information from the internet routing registry, SIGCOMM workshop on Mining network data, p.317-322, September 11-15, 2006, Pisa, Italy
5) Villamizar, C., Alaettinoglu, C., Meyer, D., and Murphy, S. 1999. Routing policy system security. IETF RFC 2725.
6) RIPE Document 201, 2001.[RRCC. http://www.ripe.net/projects/rrcc/]

7) P. Mahadevan et.al The Internet AS-Level Topology: Three Data Sources and One Definitive Metric. ACM SIGCOMM Computer Communication Review, 2006.
8) B. Zhang, R. Liu, D. Massey, and L. Zhang. Collecting the Internet AS-level topology. ACM SIGCOMM CCR, 2005.
9) G. Siganos and M. Faloutsos. Analyzing BGP Policies: Methodology and Tool. In IEEE INFOCOM, 2004.
10) G. Siganos and M. Faloutsos, ―A Blueprint for Improving the Robustness of Internet Routing, Security '06, 2006
11) K. Sriram , O. Borchert , O. Kim , P. Gleichmann , D. Montgomery, A Comparative Analysis of BGP Anomaly Detection and Robustness Algorithms, Cybersecurity Applications & Technology Conference for Homeland Security, p.25-38, March 03-04, 2009
12) http://www.iana.org/numbers/
13) http://www.ripe.net/
14) http://www.irr.net/docs/list.html
15) RADB db. ftp://ftp.radb.net/radb/dbase/.
16) RIPE db. ftp://ftp.ripe.net/ripe/dbase/.
17) IRRd. http://www.irrd.net/
18) http://www.isc.org/index.pl?/sw/IRRToolSet/
19) http://www.routeviews.org/.
20) http://www.ripe.net/projects/ris/
21) http://irl.cs.ucla.edu/topology/
22) Worldwide Infrastructure Security Report Arbor Networks 2009
23) http://asert.arbornetworks.com/2008/05/using-rpki-to-construct-validated-irr-data/
24) http://www.ietf.org/dyn/wg/charter/sidr-charter.html
25) http://irl.cs.ucla.edu/topology/