On the structure and application of BGP policy Atoms

Yehuda Afek Omer Ben-Shalom Anat Bremler-Barr

Abstract—**The notion of Internet *Policy Atoms* has been recently introduced in [1], [2] as groups of prefixes sharing a common BGP AS path at any Internet backbone router. In this paper we further research these 'Atoms'. First we offer a new method for computing the Internet *policy atoms*, and use the RIPE RIS database [6] to derive their structure. Second, we show that atoms remain stable with only about 2-3% of prefixes changing their atom membership in eight hour periods. We support the 'Atomic' nature of the *policy atoms* by showing BGP update and withdraw notifications carry updates for complete atoms in over 70% of updates, while the complete set of prefixes in an AS is carried in only 21% of updates. We track the locations where atoms are created (first different AS in the AS path going back from the common origin AS [1]) showing 86% are split between the origin AS and it's peers thus supporting the assumption that they are created by policies. Finally applying atoms to "real life" applications we achieve a modest savings in BGP updates due to the low average prefix count in the atoms.**

Keywords—**BGP, policy routing, routing protocols**

I. Introduction

The Internet today connects together thousands of independent network environments administered by different bodies called autonomous systems (AS), each governing a group of networks assigned to them commonly referred to as prefixes. Routing between ASs is governed by the Border Gateway Protocol (BGP) Version 4. The default metric used by BGP to make a decision on the next AS to pass traffic through is the number of ASs on the path to the destination (AS path). The basic entities in the BGP algorithm are therefore the ASs and the prefixes. To allow an AS to set routing policies according to financial and contractual agreements with other ASs, BGP supports overriding the default metric, per prefix, with other preferences. As a result prefixes in the same destination AS may have different AS paths though starting at the same router.

School of computer science, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: {afek,obensha,natali}@cs.tau.ac.il.

[1] The origin AS is the last AS in the AS path and is the AS that has the prefix allocated to it. In this article we also refer to it as the 'owning AS' to avoid possible confusion with an AS sending traffic to a prefix

In [1] [2] Broido and claffy suggest the existence of a third aggregation entity that may introduce another intermediate level of hierarchy to the Internet, which they called a *policy atom*. A *policy atom* groups together a number of prefixes that share the same policy. Roughly speaking a *policy atom* is a maximal group of prefixes that have the same AS path to them from any major router (i.e. a router with default-free BGP table) in the Internet. The Atom count was determined to be about 20K compared to the 120K prefixes in the Internet at the time of writing and compared to about 12K active ASs. The fact that the Atom count is closer to the AS count than it is to the prefix count raised the possibility that using the *policy atoms* instead of prefixes for some applications can achieve substantial savings in both router resources and BGP traffic. The aim of this paper is to enlarge and expand on those suggestions, validate the *policy atoms* as policy entities with 'Atomic' nature and test the ability to use them to save in Internet update traffic. In most cases in the article we will refer to the *policy atoms* simply as 'atoms' for brevity.

In this paper we have studied the following:

1. Verification of the stability of the method used to calculate the *policy atoms*: Broido and claffy used a 'snapshot' of many BGP tables to measure the structure of the *policy atoms*. However dynamic changes and concurrent BGP updates while taking the snapshots may disturb this measurement. In order to overcome such problems we used an alternate calculation method that only used information from times in which the prefixes belonging to an atom remained 'stable' (no updates for them has been seen from any source) for a duration of about 15 minutes.
2. Formation of atoms: A single AS with multiple prefixes is broken into several *policy atoms* when there are two or more known AS paths to the different prefixes of this AS in the view of one or more Internet backbone routers. We defined the AS at which atoms are created as the closest AS to the origin or owning AS, in which the AS-path to some set of prefixes in the owning AS differs from the AS-path to another set of prefixes in the same owning AS. About 85% of the atom creation points showed the atoms created between the owning AS and an AS it peers with. These

results support the hypothesis that atoms are indeed the product of the Internet policies enforced by the network administrators rather than the result of network faults.
3. Stability of atoms: Since atoms are created from the viewpoint of distributed routers and likely to be used in a distributed environment it is very important to see how consistent their structure is. Our measurements show that the atoms tend to be relatively stable with only about 2-3% of all prefixes changing atom membership in a period of 8 hours. The rate goes up to about 4-5% for a full day period and up to about 12% in a week period. This rate of change is relatively low compared to the prefix update rate.
4. Correlation of BGP updates to atom structure: The BGP protocol tries to group together in an update all prefixes which share the same attributes. To further check if the atoms are 'real' entities we checked how often complete atoms appear in BGP updates. We have found that 75% of the updates contain a complete atom in them. The matching result for complete AS prefix sets is much lower at about 20%. This indicates that most changes affect complete atoms and not complete ASs and therefor that atoms really are the base units of internet BGP routing.
5. Application of *policy atoms* for Internet routing: Finally we checked the gain that may be achieved by replacing the prefixes in the update with the atom they belong to. Using atoms in BGP updates gives on average a reduction of about 33% on the announcement size. Farther we show that the maximum average reduction attainable is about 66%. This provides an example of how atoms can enhance the BGP protocol efficiency.

The remainder of the paper is organized as follows. Section II introduces the basic definitions of *policy atoms*, and describes the two different ways in which atoms are calculated. Basic statistics for the prefix, atom and ASs entities are give in Section III-A. Section III-B deals with the stability of the Atoms. Section III-C analyzes the correlation between the Atom structure and the contents of the BGP update records. Section III-D discusses the creation of atoms and attempts to decide which AS locations are prevalent. Section IV checks the feasibility of using Atoms to reduce the communication complexity of BGP updates and finally Section V concludes this paper with a summary and some suggested topics for future research. This paper assumes the reader is familiar with the basics of BGP routing, a short introduction is available [4], [5].

II. Definitions and Calculations of *policy atoms*

The notion of BGP *policy atoms* as a higher level grouping of networks in an AS that appear with the same routing characteristics from the view point of any backbone internet router was recently introduced by Broido and claffy. They define an atom to be a group of prefixes such that for any default free router, all the prefixes in the group have the same BGP path attribute. They then performed the atom structure calculations only on prefixes appearing in all route tables. We define an Internet *policy atom* slightly differently to be a group of prefixes $\{P\}$ such that for any index i, j, prefixes $P_i, P_j \in \{P\}$ and for any router A that holds a full BGP table in the Internet, the BGP route $B_a(P_i)$ from A to P_i equals $B_a(P_j)$. Or in words, maximal group of prefixes that have the same AS path to them from any default free router in the Internet. Prefixes missing from the view point of some internet routers are a known phenomenon [8], in such cases we regarded two groups of prefixes, one seen by a specific Internet router and one that is not seen to be in different atoms even if they shared a common AS path on all routers that saw both groups [2]. To illustrate this let's consider a brief example:

```
Router X
   Network        NextHop  Path
 *> 11.11.0.0/16  9.8.7.6  121 143 11 ?
 *> 12.12.0.0/16  9.8.7.6  121 143 11 ?
 *> 13.13.0.0/16  9.8.7.6  121 143 11 ?
 *> 14.14.0.0/16  12.1.1.1 121 3 45 11 ?
```

```
Router Y
   Network        NextHop   Path
 *> 11.11.0.0/16  23.4.5.6  13 143 11 ?
 *> 12.12.0.0/16  23.4.5.6  13 143 11 ?
 *> 14.14.0.0/16  11.2.2.2  13 3333 11 ?
```

Based on the view of these two routers only we would consider AS 11 to possess three atoms:

1. Atom #1: Prefixes 11.11.0.0/16 and 12.12.0.0/16
Atom size = 2
2. Atom #2: Prefix 13.13.0.0/16
Atom size = 1, split from Atom #1 by missing path on Y
3. Atom #3: Prefix 14.14.0.0/16
Atom size = 1, split from #1 and #2 by having a different first peer in the AS paths in X (also in Y)

In [1] it is shown that information from the a relatively low number of router (8-9) sources is enough to get a very good approximation of the atom structure to within a few percent. Adding more router sources changes only slightly the atom resolution. Since the focus of this study was to check the validity and 'real life' status of the atoms rather than come by the most exact atom structure and since all our findings are based on very distinct results we concluded that using the RIPE database for 13 peers is a good enough. The less than perfect derivation of atom structure is not expected to affect our findings in any significant way. For the same reason we did not perform a num-

[2] Our calculations amounts to the same definition as the one by Broido and claffy if we only regard prefixes that exist in all route tables

ber of refinements previously deployed, such as using only crown atoms (as defined in [1]) or checking if the IP range for some atoms are included in higher levels of aggregation in other atoms. We simply note that both could have enhanced the accuracy of the calculation and reduced the total number of atoms. We also did not attempt to resolve BGP AS sets [3] and Multi Origin AS [4]. We were satisfied to accept their effect on the AS PATH alone.

We calculated the atoms in two different ways:

.1 Calculation of atoms by the 'snapshot' method

This method uses the route table information supplied in the RIPE snapshot. Using those dumped route tables we derived the atom structure according to the formal definition above, i.e., grouped prefixes that share a common AS path in all the BGP tables available. We consider the main hazard in such an approach to be the possibility that a change in the atom structure has occurred but that at the time of the snapshot not all routers received the information on the change i.e., that we may get an incorrect distributed snapshot. This risk seems real considering the fact that Internet convergence is known to take up to 15 minutes [3] and that on average 17% of all prefixes are updated in each day period [4].

.2 Calculation of atoms by the 'quiet period' method

To solve the potential problem with the 'snapshot' method and see if this problem is common we introduced a second method. We used the update records in the period of 4 hours following the time of the snapshot to track the route table during that time. We defined a prefix to be "safe" for calculations *iff* at the time of calculation there was a period of 1000 seconds around the calculation in which no update for that prefix was recorded assuming this period of time was large enough to ensure convergence. We therefore started with a snapshot of a route table as a starting point and tracked the route table along those 4 hours. At each 1000 second checkpoint we calculated the atom structure. We then defined atoms as cliques of prefixes that appeared together consistently in the same atom at each or most checkpoints [5]. Our results show that in the time period checked this second method of calculation produced results similar to within a few percent of the ones achieved using the first method.

[3]The BGP AS SET attribute is set when aggregation of paths takes place. The aggregating router places all the ASs removed from the path in a set, an AS path with an AS set may look like 14 161 [15 17 19]. This removes all the information on the AS path after the aggregation point. Having no information we simply regarded a path to be the same *iff* the whole BGP path was the same including paths with AS sets. We do not consider this to be a problem due to the relatively low number of paths with AS SET in them.

[4]MOAS was shown to be mostly due to multi homing without BGP or with private AS numbers or to be caused by short lived fault [7]. The scope of MOAS observed by us was just 5% of all prefixes

III. Results

This section discusses in detail our findings about the *policy atoms* structure, creation and validity. A possible usage of the results is shown in section IV.

A. General statistics for ASs and Atoms

In this Section we present some general statistics on the entities referred to in this paper as support for the next sections.

The average number of Atoms calculated by us was 25K. The number of Atoms is much closer to the number of ASs seen (12.5K) than to the number of prefixes seen (115K). Figure 1 provides general distribution statistics for the sizes of the ASs, Atoms and BGP updates in the period processed. It is important to note that both ASs and Atoms tend on average to include a low number of prefixes. Also notice the close match between Atom size distribution and update size distribution, the correlation between these two is discussed in detail in Section III-C.

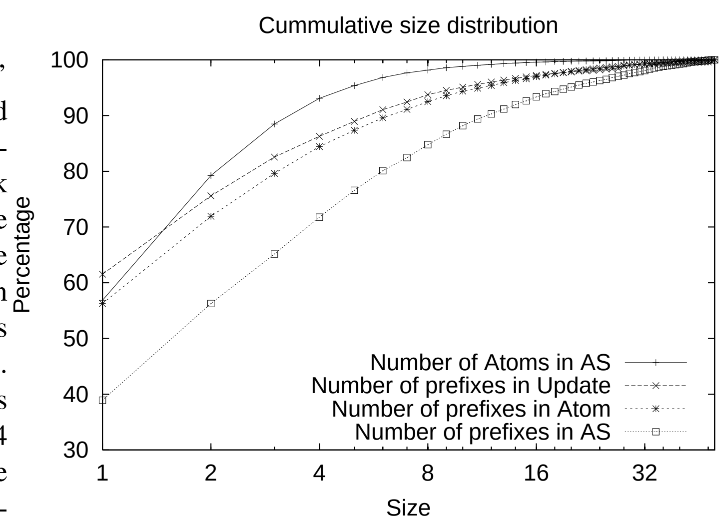

Fig. 1. Distribution of AS, Atom and updates sizes

B. Stability of policy atoms

After calculating the atom structure the next step we took is to check the stability of the atoms calculated. If the atoms are to be useful in a distributed environment than

[5]We varied the level of certainty for atom membership in the cliques from 100% (appearing together in all check points) to 50% with very little difference in the structure of the atoms or their stability

a common view of the atom structure needs to be maintained, rapid changes in the membership would make that much harder.

We compared Atoms from different periods or calculation methods in two ways: First the percent of atoms that had a complete match i.e., had the exact set of prefixes in the first and second sets. Second the percent of prefixes that remained grouped together when an atom in the first set was matched to the closest atom on the second set. For example one atom seen with 5 prefixes at one period later seen split to two atoms, one with four prefixes and one containing a single prefix would be considered 0% match for complete atoms but 80% match for prefixes since only one prefix is considered changed by leaving the main atom.

The results of comparing atoms between different periods are summarized in table I, which shows that the atoms calculated using both methods discussed are pretty stable with only about 5% of all atoms changing their exact prefix set over an 8 hour period. When considering the prefixes membership match we show only about 2-3% difference for the same time period. The numbers go higher to about 8% of atoms and 4% of prefixes changing over a day and about 20% of atoms and 14% of prefixes changing over a one week period. This means that keeping the atom membership information accurate to within 2-3% in a distributed environment may only take a few thousand updates in a few hours time window. Compared to the normal volume of internet updates during the same period of time this does not look prohibitive.

C. Correlation of atom structure to Internet update records

To use bandwidth efficiently BGP groups together in one update message information received during a set time window for a group of prefixes. This is done only if those prefixes share all BGP attributes. Since all members of an atom share the same path attribute by definition, and should share all other attributes in practice, a good correlation between the update structure and the atom structure is to be expected if atoms are really driven by policy. Having calculated the atom structure we were able to check the correlation of the atoms calculated to the prefixes seen in update records during the 4 hour period following the calculation. The study shows that there is indeed a good correlation between atom structures and the updates seen. In general the atom structure calculated by the 'quiet period' method showed a slightly better correlation on average but not by much. In both cases the number of atoms seen in their entirety in an update was about 70%-75% of the number of updates seen. We also checked the correlation between the prefixes in the update records and the full set of prefixes with the same origin AS. The results in this case were much less favorable with only about 21% of AS appearances in an update containing the whole prefix set for the AS. When checking how many different atoms were represented in each update message the results show that the vast majority of updates (86% average) contained information for members of a single atom only and 10% more contained information from members of two atoms only. This also is a good indication of the 'Atomicity' of the updates. Finally we tried to pair each two consecutive updates together which simulates very closely doubling the BGP update timer. We saw no real improvement indicating the standard BGP timer for grouping prefix notifications in a single update is probably quite good.

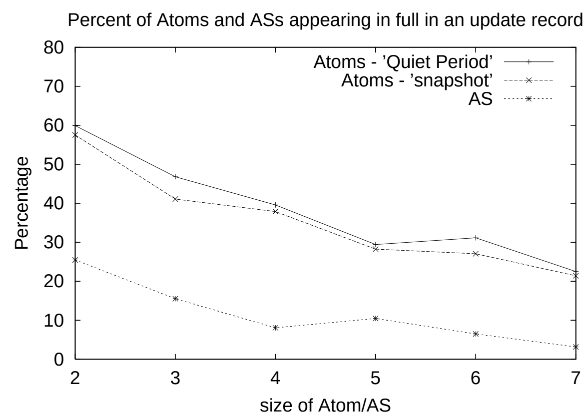

Fig. 2. Correlation level between atoms calculated using the two methods and update records

The good correlation levels seen indicate to us that the Internet infrastructure indeed changes mostly in atom units. This firmly establishes the atoms as real Internet entities.

D. Formation of policy atoms

Trying to further ascertain that the atoms are created by policy we next analyzed the location along the BGP path where atoms seem to be created. The method of calculation used included processing all AS paths recorded in all BGP tables recorded in the RIPE snapshots we retrieved. We defined the splitting point of any two atoms as the first AS going from the origin AS that is different for the two atoms. Stated differently it is the length of the minimal AS path starting from the origin AS not shared by the two atoms. The first split point is therefor the first AS going from the origin AS that is not shared by all atoms and the last split point is the shortest length of the AS path such that the partial AS path of this length is different in each atom. In the example of the routing in Section II the first split is at position 1 as two atoms are already seen at the

Time span	Quiet Atoms	Snapshot Atoms	Quiet Prefixes	Snapshot Prefixes
8 Hours	95.6 %	95.3 %	97.4%	97.7 %
1 Day	92.3 %	91.6%	97%	97 %
1 Week	78%	77.5%	88%	86%

TABLE I

STABILITY LEVELS FROM ATOMS CALCULATED USING THE QUIET TIME AND SNAPSHOT METHODS

owning AS itself. The split was due to the fact that one of the routers did not see one of the prefixes at all. Considering an AS path of length 2 in the same example all atoms are already seen using just the first 2 AS on the path starting from the origin as there are 3 different combinations already. The third atom was split by router X - prefix 14.14.0.0/16 shows a partial AS path of length two (345 11) while the other three prefixes are showing a different partial AS path (143 11). Using longer AS paths will not allow more than 3 different partial AS paths and will therefore not generate any new atoms.

To calculate the split points we grouped for each AS all the AS paths recorded in the snapshot route table. We then removed all AS sets and duplicate AS in the paths (as both do not change the AS path length in a way that changes the calculation) but we keep knowledge of AS duplication since we consider a path similar to 278 14 14 different than a path of 278 14. Even though both pass AS 278 followed by AS 14 the extra appearance of AS 14 is due to a policy and defines different atoms. Figure 3 shows a CDF graph for the location in the AS path where the first atom split is seen, for the percent of atoms created at each distance and for the location when the atoms split is full (i.e. all atoms are already seen).In the calculations we did not consider at all those atoms that show Multiple Origin AS (MOAS) conflicts. The calculation show that 85% of atoms are created between the source AS and it's immediate peers. This is consistent with the assumption that the *policy atoms* are created by policy because in the vast majority of cases the policy made is within the source AS by the customer or by a provider seeking to affect the traffic to or from the customer. This numbers sit very well with the Internet policies and relations described in [5] regarding types of relations between ASs in the Internet. We also tried to check what type of policy was involved in the creation of the atoms. Unfortunately except for AS prepending, which can only happen at the AS seen duplicated, it is quite hard to differentiate between affects of different policies on AS paths.

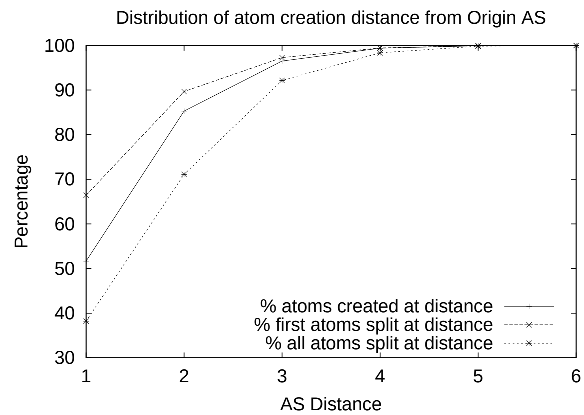

Fig. 3. Distribution of atom creation locations: first atom creation position, count of all atoms creation position and position where all atoms are already created.

IV. COMPRESSING BGP UPDATE TRAFFIC

Broido and claffy [1] predicted that atoms can be useful in lowering the size of the initial exchange between BGP routers but specified that using atoms to reduce the traffic for updates later on would be more difficult to achieve. They also estimated that the number of route announcements can be reduced by 50% based on renumbering the IP space into fewer CIDR [6] blocks based on the atom information. This is a very extreme measure as it requires a renumbering which is very unlikely to be accepted. We checked the effect of only manipulating the BGP updates. This method has the potential advantage of using both prefix based announcements and atom based announcements and avoiding the need to make radical changes in BGP or the IP allocation scheme.

As a result of the good correlation of atom structure to BGP update traffic discussed in Section III-C it is possible in most cases to compress the BGP update traffic by replacing reference to all prefixes in an update with the ID of the atom they belong to. This means, of course, that both sending and receiving router have synchronized atom table. This in turn means either some central body will need to create the atom structure and propagate it or that a distributed method of atom calculation be devised. The method of doing either is outside the scope of this paper.

[6]CIDR (Classless Inter-Domain Routing) is announcing superset or aggregate networks using a generalized network prefix size instead of being limited to the 'traditional' subnet masks of 8, 16 or 24 bits

Assuming a suitable system for the creation and distribution of atom information and roughly estimating the size of an atom record to be the same as a prefix record a compression of about 1/3 is immediately possible for the prefix section of the update.

The results shown can be enhanced by a more precise atom membership calculation as discussed in Section II and by allowing an update to contain an atom and a list of exception prefixes that are not affected by the update. In spite of this a theoretical upper bound on this compression method can be shown. Considering that no update can be compressed to less than a single entry we note that the number of updates was consistently about 1/3 of the total number of prefixes contained in the updates. Therefore the achievable savings would range between the 30% we have shown above and an upper bound of 66%. To achieve further reduction in update traffic will mean changing the way BGP handles updates [10]. We can also suggest that the very knowledge of the Atom structure may be useful in filtering out 'noise' from the updates. This can be done, for example, by demanding a longer update timer on updates that show change for only some of the prefixes belonging to an atom (since we expect updates on full atoms) or shortening it when information for a full atom is received. This can be the subject of future research.

V. Conclusion

In this paper we have studied the characteristics and implementation of the *policy atoms* introduced by Broido and claffy, we have shown an alternate method for calculating the atom structure, proved the validity of the *policy atom* as both an 'atomic' unit and a result of policy and for the first time studied the applicability of the *policy atom* entities. Our results show that the atoms are real entities closely tied to the observed data from Internet routing traces and also suggest that atoms can be used in some cases to achieve savings in bandwidth in Internet routing updates.

Looking ahead we see a number of important issues that need further research if atoms are to take a substantial part in Internet measurements and operation, chief among these issues are the following two topics:

1. Getting the knowledge of the atom structure proliferated to all Internet routers.

This can be done by a central body performing the calculation and distributing the results similar to our way of calculation, by getting the origin AS to tag its prefix atom membership using BGP communities based on knowledge of its policies or by devising a suitable distributed algorithm that will allow the Internet backbone routers to calculate their structure separately.

2. Better network faults handling

Network faults normally affect whole atoms when the fault occurs outside the bounds of the owning AS. This is strongly supported by the fact that BGP updates contain full atoms in most cases. Knowing the atom structure may allow a better understanding of the scope of a fault and thus support a more efficient reaction to the fault. One way to use the information can be to use different timers for announcements including full and partial atoms.

Another question directly tied to this is the question of the actual physical paths traversed by Internet traffic i.e. do destinations in the same prefix and atom pass the same router path as well as the same AS path.

Acknowledgments

We thank Andre Broido from CAIDA and Edith Cohen from AT&T for helpful discussions and Jennifer Rexford from the IMW steering committee and the referees for very helpful comments.

References

[1] A. Broido, kc claffy, Cooperative Association for Internet Data Analysis - CAIDA, San Diego Supercomputer Center, University of California, San Diego. "Analysis of RouteViews BGP data: policy atoms". In *Proceedings of NRDM workshop Santa Barbara*, May 2001

[2] A. Broido, kc claffy, Cooperative Association for Internet Data Analysis - CAIDA, San Diego Supercomputer Center, University of California, San Diego. "Complexity of global routing policies", In *Proc. IMA Special Workshop: Mathematical Opportunities in Large-Scale Network Dynamics*, August 2001

[3] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanianitz, "Delayed internet routing convergence", In *Proc. ACM SIGCOMM*, September 2000

[4] Timothy G. Griffin, "Tutorial: An Introduction to Interdomain Routing and BGP", In *Proc. ACM SIGCOMM*, August 2001

[5] "Supplementary data for the BGP atoms extended abstract", `http://www.tau.ac.il/~obensha/BGP_background.ps`

[6] RIPE Routing Information Service, `http://data.ris.ripe.net/`

[7] Xiaoliang Zhao, Dan Pei, Lan Wang, Dan Massey, Allison Mankin, S. Felix Wu,Lixia Zhang, "An analysis of BGP multiple origin AS (MOAS) Conflicts" In *Proc. ACM SIGCOMM Internet Measurement Workshop*, 2001

[8] Craig Labovitz, Abha Ahuja and Michael Bailey, "Shining Light on Dark Address Space" Tech Report, Arbor networks , June 2001

[9] Complete article code `http://www.math.tau.ac.il/~obensha/thesis/thesis_code.`

[10] Dan Pei et al, "Improving BGP Convergence Through Consistency Assertion" In *Proc. IEEE INFOCOM*, 2002