

A Measurement Study on the Distribution Disparity of BGP Instabilities

Meng Chen*, Mingwei Xu*, Yuan Yang*, and Qing Li**

*Tsinghua National Laboratory for Information Science and Technology

*Dept. of Computer Science and Technology, Tsinghua University, China

**Graduate School at Shenzhen, Tsinghua University, China

chenm11@mails.tsinghua.edu.cn xumw@tsinghua.edu.cn

yyang@csnet1.cs.tsinghua.edu.cn li.qing@sz.tsinghua.edu.cn

Abstract—BGP measurement is important for monitoring and understanding the Internet anomalies. Most of the previous works on BGP measurement rely on aggregated statistics from BGP monitors, e.g., total updates. However, BGP events may have quite limited visibility. Therefore, merely investigating aggregated data may lead to misunderstanding Internet instability, e.g., overestimating the impact of monitor-local events. In this empirical study, we demonstrate how BGP data are distributed among a large number of monitors. We define eleven features as the analysis targets, and three metrics to quantify disparity. We apply the method to 1.14 TB data and find that the distribution of most of the features is quite uneven, and different types of feature illustrate different levels of disparity. We also observe long periods of persistent high disparity, and a small set of cross-feature highly active monitors. Our analysis highlights the necessity of per-monitor data analysis in future BGP measurement study.

I. INTRODUCTION

Border Gateway Protocol (BGP) is the de facto inter-domain routing protocol, and BGP dynamics is an important indicator of the stability and healthiness of the Internet. There is extensive literature on measuring and investigating BGP traces. By analyzing the changes in BGP updates and tables, many researchers seek to detect and understand the impact of Internet disruptive events such as cable cuts, large-scale attacks, blackouts, and configuration errors [1] [2].

We find that most of the works on BGP measurement leverage aggregated data from BGP monitors, instead of investigating data from each monitor separately. For example, some works analyze the total quantity of updates [3] or path changes [4], and adopt machine-learning or statistical techniques to detect anomalies. However, while such methods save much time in data-processing, they may lead to potential artifacts and misunderstanding BGP instabilities, e.g., over-evaluating the impact of monitor-local events, which are visible to only a rather limited area of the Internet.

The potential artifact is essentially rooted in the principle of BGP, which largely confines the propagation scale of update. Firstly, BGP is an incremental protocol: after the initial exchange of routing information, BGP-speaking routers advertise only changes to that information. Therefore, an update message cannot propagate far unless it changes most of the best routes in the routers it traverses. Secondly, routing policy

may limit the propagation of certain updates. For instance, an Internet Service Provider (ISP) usually does not advertise the routes learned from peers to its providers and other peers; hence the updates for those routes will not be propagated in these directions.

However, such knowledge could be vague; without real data, it is still largely unknown how different types of BGP instabilities are actually distributed among the hundreds of public BGP monitors deployed in the world. Lacking this information conceals the pervasiveness of monitor-local events, and may lead to underrating the necessity of per-monitor measurement methods. The purpose of this paper is to investigate the distribution disparity of BGP instabilities, with large data.

We use eleven features as the analysis targets, including ten basic features and the update for Highly Active Prefixes (HAP). The basic features are update patterns extracted from BGP traces, and each feature indicates a particular aspect of instability, e.g., successive duplicate announcements to the same prefix indicate pathological BGP behavior. These basic features (with small differences) are used in previous research to assess the stability and healthiness of the BGP system [2] [5] [6]. On the other hand, while Highly Active Prefixes (HAP) are believed to be an important indicator of BGP instability [7], the traditional definition of HAP is based on a simple count of total update, and it remains unknown how these updates are distributed, and whether an HAP is due to local events.

To quantify distribution disparity, we use three metrics, including Participating Monitor Ratio (PMR), Gini Coefficient (GC), and Concentration Ratio (CR). These metrics focus on different aspects: PMR records the ratio of monitors observing a certain feature; GC quantifies the overall disparity level; CR measures the contribution of the top few monitors. The combination of the three metrics helps providing a comprehensive analysis result.

We apply the method to six months' (May to October 2013) BGP updates; the total data are 1.14 TB. We conduct a multi-faceted analysis of the data and discuss the implications of the measurement results. Apart from the basic information about the distribution of the features, we also analyze the time pattern of disparity, as well as a detailed analysis of some highly active monitors. Our major observations are concluded as follows.

- 1) The distribution of most of the basic features is rather unbalanced. In particular, the features indicating pathological behaviors and policy changes have much higher disparity than the others. The result implies that pathological BGP behaviors and policy changes are highly localized, and data aggregation is most unsuitable in such cases.
- 2) For a large portion of the basic features, distribution disparity is in positive correlation with the total quantity of the features. Since many works use high total quantity as an important indicator of anomaly, we highlight the potential risk of such a practice: the high total quantity is probably due to the sudden rise in the contribution of few monitors.
- 3) We observe persistent high-disparity periods for most features, and the longest period lasts for more than 20 days. We suggest special attention being paid to such phenomena in BGP measurement that relies on aggregated data when, for example, per-monitor analysis is infeasible due to limited time and resource.
- 4) A large portion of HAPs are caused by local events. We suggest adopting multiple dimensions (e.g., consider both total update quantity and the PMR value of these updates) in determining HAPs.
- 5) There is a small set of BGP monitors that are persistently much more active (i.e., observe high instability) than other monitors, even across multiple features. This observation further highlights the necessity of per-monitor data analysis, and more sophisticated data pre-processing technique if aggregation is conducted.

The rest of the paper is organized as follows. Section II presents the related works. In Section III, III-A describes the features, III-B defines the metrics, and the other part of the section introduces data collection and pre-processing. Section IV and V illustrate the results for the basic features and the HAP-update feature respectively. Section VI demonstrates the analysis of the most active monitors. Finally in Section VII, we conclude the paper.

II. RELATED WORK

There are many publications on BGP measurement, in this part we discuss a subset of them. Most of these works are about anomaly detection and analysis, and a common practice is analyzing aggregated data. Besides, some works use a small number of monitors; they are also prone to localized events. We know few works that propose per-monitor analysis methods.

BGP data aggregation. Labovitz et al. [5] raised a set of update patterns to measure the healthiness of the inter-domain routing. These patterns are based on successive BGP update types; for example, duplicate updates indicate pathological behaviors. Ten years later, Li et al. improved the pattern categorization method [6]; and in a later work Li and Brooks used these patterns to identify the impact of some famous Internet disruptive events [2]. In these works, data are aggregated for

analysis. The features defined in this paper are largely based on the patterns defined in these works.

Oliveira et al. adopted a simple count of total update to determine Highly Active Prefixes [7]. However, detailed distribution of the updates for HAPs is not mentioned. This paper seeks to analyze this distribution and discuss its impact.

Liang and Zhang proposed a statistical technique based on pattern-matching to identify and locate familiar BGP routing instabilities [1]. Only 10 monitors were used, and data from these monitors were aggregated before applying the technique. In another work, Mai et al. raised a two-phase anomaly detection method [3], in which updates are separated by the origin Autonomous System (AS) of IP prefixes instead of by monitors. In this way, BGP withdrawals (no AS-path information) are excluded from analysis, and like many other works in this field, the impact of monitor-local events cannot be eliminated. Another example of data aggregation is the work by Menon and Pottenger, in which a higher-order collective classifier is proposed to detect and classify Internet events [8].

Few used monitors. Deshpande et al. applied a statistical change-detection technique to the data from only four BGP monitors to detect BGP anomalies [9]. Potential artifact due to monitor-local dynamics could be strong, and the small number of monitors seems deficient in providing a comprehensive picture of inter-domain instability. Zhang et al. utilized a set of fixed patterns to search and identify routing anomalies [10]. Although per-monitor data are analyzed, the number of used monitors is less than 10, providing a rather limited view of the Internet.

Per-monitor analysis. While the importance of route monitor selection has been recognized [11], there has been little effort in developing per-monitor analysis techniques. In a more recent work, Comarella and Crovella focused on large-scale coordinated rerouting events reflected in daily routing table changes [12]. They adopted a tensor model, and the three dimensions are AS, prefix, and time. The elements of the tensor record path change information. By factorizing the tensor, they detected large BGP re-routing events that had never been reported before. In the work by Chen et al. [13], Update Visibility Matrix (UVM) is used to record whether the update for a prefix is observed by certain monitor ASes. In this way, monitor-local event (shown as a narrow submatrix in UVM) could be explicitly filtered out. They detected many Internet-scale events that had remained unreported.

III. METHODOLOGY

A. Features

We analyze two types of features. The first type (called the **basic features**) consists of ten different BGP trace patterns, listed in TABLE 1. Some of these features reflect normal routing changes, and some indicate pathological behaviors. The basic features are the major analysis targets. These features have been used by Labovitz et al. [5] and Li et al. [6] [2] to assess the healthiness of inter-domain routing of the Internet.

The second type is the BGP updates for Highly Active Prefixes (HAPs), denoted as **HAP-update**. We define HAP

Feature	Description
U	BGP update, direct indicator of inter-domain instability.
A	BGP announcement.
W	BGP withdrawal.
AW	Withdrawal after announcement to the same IP prefix; could be normal or pathological.
WWDup	Duplicate withdrawals to the same IP prefix; pathological.
WADup	Duplicate announcement after withdrawal to the same IP prefix; due to transient topological changes, policy changes, or pathological oscillation.
WADiff	Announcement (different AS path or next hop) after explicit withdrawal to the same IP prefix.
AADup1	Duplicate (identical) announcements to the same IP prefix; pathological because BGP is not supposed to send repeated announcements.
AADup2	Duplicate (same AS path and next hop, but not identical) announcements to the same IP prefix; mainly due to routing policy changes.
AADiff	New-path announcement after announcement to the same IP prefix; implicit withdrawal due to route changes.

TABLE 1: Basic features.

in the same way as [7]. That is, if the number of updates for an IP prefix P is higher than a pre-defined threshold, P is an HAP. We decide the threshold as follows.

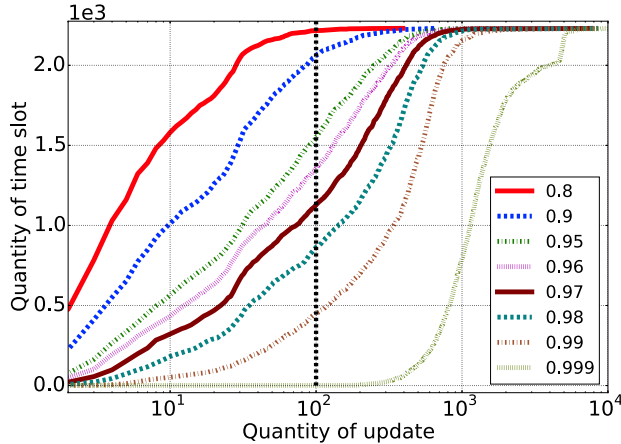


Fig. 1: The number of time slot satisfying different update quantity ranges.

In every 20-minute time slot during October 2013 (we explain time slot and the analyzed period in Section III-E), we record the quantity of updates for each IP prefix. In Fig. 1, we illustrate the number of slots y (y axis) with ratio r (curve tag) of their IP prefixes having the quantity of updates equal to or lower than x (x axis). We set the threshold, which can be represented as a vertical line, such that it intersects the 0.8 and 0.999 curves at very high and very low values respectively, and intersects the other curves at moderate values. This is based on the assumption that most slots have a quite small ratio (but non-zero) of HAPs, and a rather small number of slots have a high ratio of HAPs. In this paper, we set the threshold to 100, as indicated by the vertical dashed line in the figure.

B. Metrics

Participating Monitor Ratio (PMR): the ratio of monitors that observe any feature data within a given time period. For example, in a time slot, 80 of 123 monitors observed one or more BGP withdrawals, then $PMR(W) = 0.65$. It is the major metric to distinguish between localized events and Internet-scale events in the work by Chen et al. [13].

Gini Coefficient (GC): a widely used indicator in the field of economics to assess data dispersion. As shown in Fig. 2, the curved line is the cumulative share of the monitors from lowest to highest. The line at 45 degrees indicates perfect equality. The Gini Coefficient is the ratio of the area between the 45-degree line and the distribution curve over the total area under the 45-degree line, i.e., $GC = A/(A + B)$.

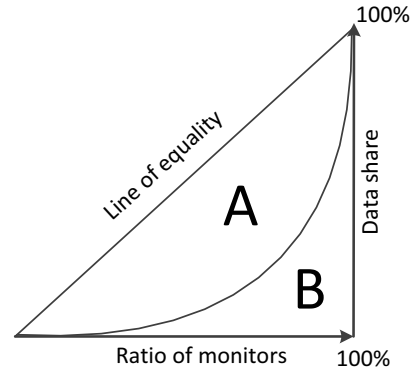


Fig. 2: Cumulative share of monitors from lowest to highest.

The value of GC is within the range of $[0, 1]$. A higher GC value reflects stronger disparity, and the value 1 means absolute disparity, i.e., all data are observed by only one monitor. The value 0 means perfect equality. To facilitate the elaboration, we categorize GC values into several levels as suggested in [14] and [15]: > 0.5 (high disparity), $[0.3, 0.5]$ (medium disparity), and < 0.3 (low disparity).

Concentration Ratio (CR): the ratio of data that are observed by the top N or $R\%$ monitors within a given time period. CR value reflects the extent of the contribution of the most active monitors. We use two classes of CR: CR_1 , CR_4 , and CR_8 record the contribution of the top 1, 4, and 8 monitors; $CR_{0.1}$, $CR_{0.2}$, and $CR_{0.3}$ record the contribution of the top 10%, 20%, and 30% monitors respectively.

C. Download and preprocess data

We download six months' BGP updates (from May to October 2013) from RouteViews [16] and RIPE [17]; both of the organizations operate multiple route collectors, and each collector has BGP sessions with multiple BGP-speaking routers (i.e., the monitors) in the Internet. We use totally 17 collectors. The RouteViews collectors are: route-views2, route-views4, route-views.eqix, route-views.isc, route-views.linx, and route-views.wide; the RIS collectors are: rrc00, rrc01, rrc03, rrc04, rrc05, rrc10, rrc11, rrc12, rrc13, rrc14, and rrc15.

The data are from 452 monitors, and add up to 1.14 TB. some of the monitors are unsuitable for our experiment thus are omitted in our analysis. a) Partial view. If a monitor has only a partial view of the Internet, it may miss some Internet-level BGP events. So we select only the monitors with global view. To achieve this goal, we set a benchmark as the quantity of prefixes in the routing table of a core router [18]. The quantity of prefixes a global-view monitor can see should be at least 90% of the benchmark. This criteria removes 269 monitors. b) Duplicate sessions. Some BGP routers have sessions with more than one collector, hence are interpreted as multiple monitors, and this may lead to the impact observed by those routers being incorrectly amplified. To deal with the issue, we select only one session among the multiple ones with the same router. This step removes 52 monitors. c) Unbalanced distribution. Some ASes own multiple monitors, which may result in the impact of events local to these ASes being overstated. So we select only one monitor in each AS. This step further removes 8 monitors. In summary, among the 452 monitors, 123 are suitable for our measurement.

Sometimes the BGP session between a route collector and a route monitor is re-established. In this situation, the complete BGP table of the monitor is sent to the collector in the form of a large amount of updates. While these updates occupy notable computing and storage resource, they provide little information about global Internet status. In fact, BGP session reset is a special type of local event. We delete the updates caused by BGP session reset through applying the method raised by Cheng et al. [19].

D. The distribution of the monitor ASes

1) *Tier.*: We get the tier of each monitor AS by using the method proposed by Oliveira et al [20], which adopts the size of customer cone as the metric. An AS's customer cone is the ASes that are reachable through AS-level provider-to-customer links. The size of customer cone indicates the influence of an AS. The method is: the ASes with less than 5 downstream customers are stubs; the ASes with between 5 and 50 downstream customers are small ISPs (tier-3); the remaining non-tier-1 ASes are large ISPs (tier-2); the list of tier-1 ISPs are directly obtained from a website [21]. We get customer cone data from a public repository [22], which is maintained by RIPE RIS. The result for the 123 monitors is, tier-1: 8, tier-2: 37, tier-3: 43, stub: 35.

2) *Geographical location.*: Next, we map the ASes to countries/regions by using the geographic data from a website [23]. For a tier-1 AS, we record its region as 'Global'. We find that the 123 monitors are spread across 26 countries/regions in the world. The result is as follows. USA:20, GBR:13, RUS:10, DEU:9, CHE:9, BRA:9, Global:8, NLD:7, CAN:7, AUS:5, JPN:4, ITA:4, AUT:4, NOR:2, HKG:1, FIN:1, ROU:1, ARM:1, ZAF: 1, VNM:1, CZE:1, CYP:1, AZE:1, IRL:1, UKR:1, FRA:1.

In summary, the monitor ASes are of various types and are widely scattered in the Internet, which enhances the solidness of our observations.

E. The length of time slot

As stated previously, we divide the downloaded updates into 20-minute time slots to conduct the analysis. The choice of 20 minutes as the interval is an engineering decision based on the assumption that most re-routing events converge in several minutes [24]. 20 minutes are neither too short nor too long: while the decision provides fine granularity, each slot is supposed to be able to capture the complete procedure of most rerouting events within it.

IV. DISTRIBUTION DISPARITY OF THE BASIC FEATURES

The cumulative distribution (from May to October, 2013) of six metrics for the basic features are demonstrated in Fig. 3. In each sub-figure, the y-axis represents the number of time slot, and the x-axis represents metric value; each curve represents the cumulative distribution of the corresponding metric for a certain basic feature.

It is notable from Fig. 3 that among the basic features, the distribution disparity of WWDup and AADup1 is the highest. Specifically, the GC values of WWDup and AADup1 are all above 0.95 and 0.89 respectively; both are extremely high. Besides, while the PMR values of the other basic features are mostly above 0.8, those of WWDup and AADup1 are much lower: below 0.15 and 0.38 respectively. The rather low PMR values indicate that most data are observed by a quite small portion of monitors. Actually, as illustrated in the CR figures, in every slot all the WWDup data are contributed by the top < 10% monitors.

As described in TABLE 1, both WWDup and AADup1 indicate pathological behaviors. The result implies that *the impact area of pathological inter-domain routing events is usually highly confined.*

The second level of disparity includes AADup2 and WADup, according to Fig. 3a. For AADup2 and WADup, all the GC values are above 0.46 and 0.43 respectively. However in Fig. 3b and 3c, the disparity of AADup2 and WADup are not as significant, implying that in contrast to WWDup and AADup1, the contributing monitors of AADup2 and WADup are much more and the contribution share of the top monitors are not so significant.

An interesting question is: whether the difference in distribution disparity is related to the difference in total quantity? To answer it, we plot the cumulative distribution of the total quantity of the basic features in Fig. 4.

As shown in the figure, the total values of WWDup and AADup1 are the lowest, magnitudes lower than the other basic features. There are even four slots in which no WWDup occurred at all; we omit them from analysis. The result is: if the total quantity is very low, it is unlikely that the data are observed and evenly distributed at many monitors. But if the total quantity is not too small, the correlation does not exist any more. For example, while the distribution disparity of AADup2 is larger than that of A, W, AW, and WAdiff, the total number of AADup2 is higher than the total number of these features. Moreover, although the disparity of W and AW is the lowest, the total quantities of these features are not the highest.

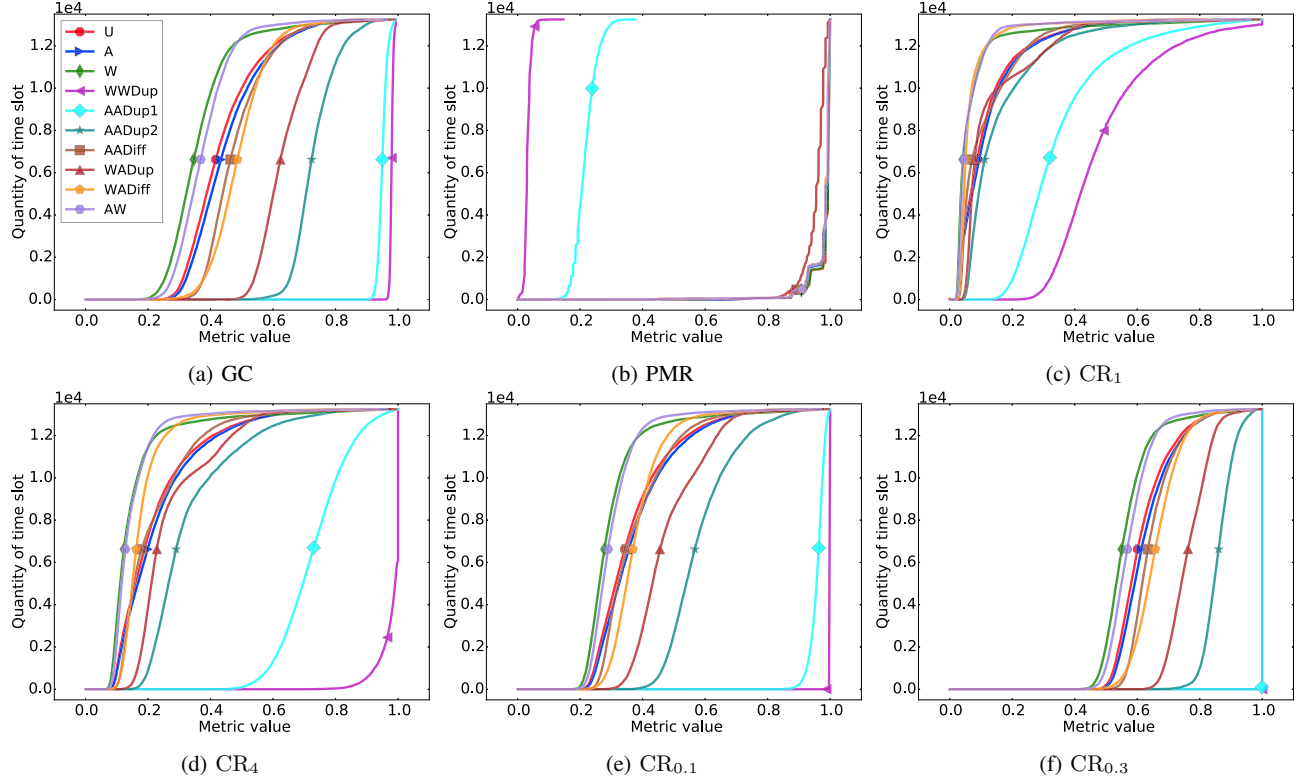


Fig. 3: Cumulative distribution of six metrics.

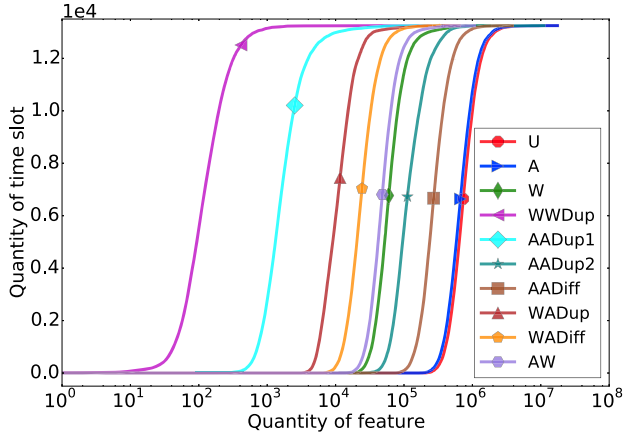


Fig. 4: Cumulative distribution of total quantity.

To summarize, except for WWDup and AADup1, we do not observe a significant correlation between distribution disparity and total quantity, indicating that very low total quantity is just one of the reasons for high disparity.

A. Correlation between disparity and total quantity

Previously, we analyzed the correlation between overall quantity and overall disparity. In this part, we investigate the correlation between disparity and total quantity in every time

slot. In other words, we check whether disparity increases with the increase of total quantity.

Due to space limitation, we demonstrate only the most representative examples, shown in Fig. 5. X-axis represents total quantity and y-axis represents metric value; every slot is represented by a dot. For a specific feature, the results of all metrics are similar; here we investigate GC.

We observe notable positive correlation between quantity and disparity for AADup2 (shown in Fig. 5a) and AADup1. Note that AADup1 indicates pathological behaviors and AADup2 indicates policy changes, both of which are of special interest for network management purposes. The result implies that *a sharp rise of AADup1 and AADup2 is likely due to the increased contribution of a very small number of monitors*. The observation is important because some previous works use high total quantity as a major indicator for anomaly [25]; we suggest more sophisticated methods that take per-monitor data into consideration because higher total quantity may indicate higher disparity, which may result in stronger artifact.

The positive correlation seems weak for U (shown in Fig. 5b) and A; it becomes stronger only when the quantity is quite large. On the other hand, there is no visible correlation for WADiff (shown in fig. 5c), WWDup, and AADiff. Especially noteworthy is that although WWDup and AADup1 are similar in rather high overall disparity, the level of disparity-quantity correlation is quite different: the latter presents a much stronger positive correlation.