Android vs. IOS: a comparative analysis over mobile operator infrastructures based on crowdsourced mobile dataset

Engin Zeydan

Accepted: 24 July 2021 / Published online: 20 August 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract
User equipment (UE)’s operating system (OS) and category types are important factors that are affecting the end-user performance in a given mobile network operator (MNO)’s infrastructure. For this reason, fair and statistically accurate observed network performance differences of UE’s OSs based on category types, MNOs or locations can be of interest for mobile telecommunication ecosystem players. This paper’s focus is on performance comparisons of UE OSs (including Android, IOS (iPhone Operating System) and Windows phones) over different UE categories, MNOs and locations based on previously collected end-to-end nationwide crowd-sourced data measurements in Turkey. The analysis results performed in this paper uses statistical comparisons of unpaired observations due to imbalance between number of observations between all OSs and yield insight on how the mobile OS types’ network performances differ using some important Key Performance Indicators (KPIs) such as downlink (DL) speed, latency, jitter and packet loss (PL). The outcome of the analysis indicate that Android devices perform better in terms of DL speed among all MNOs, whereas IOS devices are better in terms of latency values. On the other hand depending on the UE category, the performances of MNOs may vary when IOS and Android OSs are compared based on different KPIs. Additionally, IOS has shown better performance than Android in large geographical areas of Turkey. Finally, the business aspects of performing the proposed statistical OS comparisons from the perspectives of OS developers, MNOs, device manufacturers, and end-users are highlighted.

Keywords Large scale measurements · MNO · UE category · UE OS · Statistics

1 Introduction

The emerging wireless cellular technologies and vast amount of data that are generated in large-scale networks have given an unprecedented opportunity to mobile network operators (MNOs) to discover hidden values, explore new values and study characteristics of wireless devices and network infrastructures [1]. In order to observe network performance and obtain competitive advantages for managing services and networks, this generated data needs to be analyzed and evaluated [2,3]. In the telecommunication world, MNOs are leveraging data analytics approaches to harness the rapidly increasing complexity of networks. In future networks, it will be almost impossible to run a network without data driven analytics due to large number of connected User Equipments (UE) into the network infrastructure. Hence, data-driven analysis coupled with heterogeneous dataset (e.g. either publicly or proprietary crowd-based dataset) can empower automating network operations and infer some insight for additional sources of revenues (e.g. determining the set of UE operating systems (OSs) that may work well with the infrastructure based on context information).

Many wide range of research studies in different domains that attempt to understand performance of UE OSs and their comparisons are done in the literature. Some examples of those domains are privacy and security aspects [4–8], usability scores, user behaviour analysis and comparative OS evolution steps [9–11], application release frequency, important features or number of bug comparisons [12–15], mobile streaming and cellular network connectivity or control/data layer performances [16–18], impact on social media [19] or image processing capabilities [20]). In terms of mobile OS comparisons, connection management and handover capabilities of three major operating systems OSs namely Windows, IOS (iPhone Operating System) and Android are studied...
in [17]. Regarding large-scale analysis of collected Key Performance Indicators (KPIs), the authors in [21] have investigated various reasons for Packet Loss (PL) in network infrastructure such as Radio Access Technology (RAT) changes, temporary loss of service, connection termination using large-scale PL measurements for more than half a year in geographically diverse locations. Some other related large scale KPIs measurement campaigns include 3G/4G Channel Quality Indicators (CQIs) measurements in [22]. Long Term Evolution (LTE) Physical Downlink Control Channel (PDCCH) information collection in [23], Round Trip Time (RTT) measurements in [24], latency comparisons of two MNOs (TIM (Italy) and H3G (Sweden)) using 500M network latency measurements in [25], web performance of 11 different commercial mobile carriers in [26] and cellular network traffic volume over spatio-temporal dimensions in [27, 28].

In the literature, no prior efforts have concentrated on comprehensive understanding of the statistical KPI performance differences of UE OSs based on UE categories and utilized MNOs using large-scale and nationwide collected real data measurements over a long time duration in Turkey. Present works consider factors such as handover performances, connection establishment delays, network interface usage and selection capabilities in [17], the effect of mobility in [21], computer vision application performance (e.g. using OpenCV library functions) in [20], the effects of manipulating smartphone data on both Android and iOS platforms in [5], security/privacy feature comparisons in [7], mobile OS user differences in [29], review of mobile OS applications in [30], mobile app development comparisons of Android and iOS OSs in [31], the engagement levels of IOS and Android application developers in [32], application release practice comparisons of Android and IOS in [33], mobile OS platform features or ecology comparisons in [15] when designing comparative methods between existing mobile OSs. These various techniques/methods are experimental-based [17, 21], features-based [15], social-impact based [19] or image processing capability based [20] evaluations. Some benchmarking tools are also available to compare different mobile OSs and devices based on metrics such as firmware, CPU, memory, etc. [34, 35].

In this paper, answers to the following questions are investigated: (i) what kinds of differences between OSs exist? (ii) which differences between OSs are statistically significant? (iii) what are the consequences of the differences to user experience and business? Different than previous works, in this work data analysis over the network test speed data-set collected over 25 months period is performed. The results give insights into end-to-end network performance of OSs for each MNO and comprehensive understanding of the statistical KPI differences (using confidence interval (CI) evaluations) of OSs based on UE categories. Previous works in [36, 37] worked on crowdsourced dataset that are related mostly to KPI performance differences of major MNOs in Turkey over a dataset of 18 months duration. In previous related paper [36], the main focus is on performance comparisons of MNOs based on the experimental dataset that are collected apriori to compare the performance of MNOs. In another related previous work of Zeydan and Yildirim [37], reliability performance comparisons of MNOs are investigated. In this paper, these analysis are extended for comparisons of UE OSs over different UE categories for each MNO using the extended version of the dataset in [36, 37].

The main contributions of the paper can be summarized as follows: (i) Crowdsourced dataset is used to evaluate the performance differences of existing mobile OSs especially focusing on two mostly used mobile OSs, namely Android and IOS, in the world. (ii) OS-level performances are compared using KPI values such as Download (DL) speed, latency, jitter and PL with respect to different UE categories and MNOs using statistical comparison methods that are based on CI calculations. (iii) The results indicate that Android OS works better in DL speed, whereas IOS OS phones are better in terms of latency performance. In terms of PL and jitter, the performances vary depending on the MNOs. Moreover, the analysis results indicated that among all OSs, Windows OS phones perform poor in all considered KPIs. (iv) To better understand geographical comparisons of different regions at large scale, CI values are also visualized for 4G cellular networks in all cities of Turkey. (v) Finally, the business aspects of performing this kind of OS comparisons from the perspectives of OS developers, vendors, MNOs, device manufacturers, and end-users are highlighted.

The rest of the paper is organized as follows: In Sect. 2, comparative analytical results using CIs for large scale and proportions are given. In Sect. 3, performance evaluations of the comparisons of OSs based on UE categories, MNOs and geographical regions are provided while also highlighting the main takeaways, business consequences of such comparisons and possible dataset extensions. Finally in Sect. 4, conclusions of the paper are provided. Additionally, Table 1 provides all symbols and their corresponding definitions used throughout the paper.

### 2 Comparative confidence interval analysis

#### 2.1 System model and utilized KPI definitions

The statistical model used for experimental comparisons are based on the method utilized in [36, 37]. In this paper, CI values for DL speed, latency, jitter and PL for performance comparisons of phone OSs are calculated. It is assumed that there are $K$ OSs and $L$ UE categories used in a single MNO and in our system there are $M$ MNOs in total. The MNO set is denoted as $\mathcal{M} = \{1, 2, \ldots, M\}$, UE category set as
### Table 1 Symbols and their corresponding definitions used throughout the paper

| Symbol | Meaning |
|--------|---------|
| $K, \mathcal{K}$ | Number of OSs, OS set |
| $L, \mathcal{L}$ | UE categories, UE categories set |
| $M, \mathcal{M}$ | Number of MNOs, MNO set |
| $\bar{x}_k(m, l)$ | Sample mean for MNO-m, UE category-l and OS-k |
| $s_k(m, l)$ | Sample standard deviation for MNO-m, UE category-l and OS-k |
| $d_{k,1,2}(m, l)$ | Mean difference between two observations of OS-k1 and OS-k2 for MNO-m and UE category-l |
| $s_{k,1,2}(m, l)$ | Standard deviation of the mean difference between two observations of OS-k1 and OS-k2 for MNO-m and UE category-l |
| $\alpha$ | Significance level |
| $\bar{z}_{1-\frac{\alpha}{2}}$ | $(1 - \frac{\alpha}{2})$-quantile of $\mathcal{N}(0, 1)$ |
| $\mathcal{N}(0, 1)$ | Normal distribution with zero mean and unit variance |
| $s_k(m, l)$ | Sample standard deviation for MNO-m, UE category-l and OS-k |
| $n_k(m, l)$ | Total number of observations for OS-k, MNO-m and UE category-l |
| $\mu_k(m, l)$ | Mean value over $(m, l)$ observations for OS-k, m-th MNO and l-th UE category |
| $\bar{\mu}_k(m, l)$ | The standard error for the mean for unpaired comparison scenarios is used to make fair comparison of OSs in the existing dataset for fair comparison purposes. The details of the data collection architecture can be found in [37]. Note that more additional features of the UEs such as the RF characteristics of the UE environment, the firmware version of the UE, and the RF chip-set can help out to draw better comparison of OSs. |
| $\sigma_k(m, l)$ | Standard deviation over $(m, l)$ observations for OS-k, m-th MNO and l-th UE category |
| $\bar{\sigma}_k(m, l)$ | Coefficient of Variation over $(m, l)$ observations for OS-k, m-th MNO and l-th UE category |
| $p_{k1}(m, l)$ | Packet loss probability and average packet loss probability for $(m, l)$ observations for OS-k1, m-th MNO and l-th UE category |
| $\bar{p}_{k1}(m, l)$ | The standard error of observation for OS-k1 for a given m-th MNO and l-th UE category |
| $N_{k1}(m, l)$ | Number of observations for OS-k1, m-th MNO and l-th UE category |
| $C_k(m, l)$ | Confidence interval value for OS-k, m-th MNO and l-th UE category |
| $CI_{k,1,2}(m, l)$ | Confidence interval difference value for OS-k1, OS-k2, m-th MNO and l-th UE category |

### Table 2 Nationwide network performance test data statistics in Turkey

| # of measurements | 5,507,750 | # of measurements (IOS) | 1,927,784 |
| # of measurements (Android) | 3,536,595 | # of measurements (Windows) | 21,789 |
| Obser. duration | 25 months | Total # of UEs | 1,051,734 |
| M (# of MNOs) | 3 | $\alpha$, Significance level | 10 |
| K (# of OSs) | 3 | L (# of UE Cats.) | 8 |

$\mathcal{L} = \{1, 2, \ldots, L\}$ and the OS set as $\mathcal{K} = \{1, 2, \ldots, K\}$. The main aim of this paper is on comparative performance differences of OSs based on different UE categories and MNOs that are operating nationwide in different regions or cities of the country using the crowd-sourced dataset that contains measurements of UEs to test their network performance. Note that the number of measurements on different UE categories and MNOs are different. For this reason, confidence interval (CI) for the mean for unpaired comparison scenarios is used to make fair comparison of OSs in the existing dataset for fair comparison purposes. The details of the data collection architecture can be found in [37]. Note that more additional features of the UEs such as the RF characteristics of the UE environment, the firmware version of the UE, and the RF chip-set can help out to draw better comparison of OSs. Unfortunately this is not the case in our experiments due to unavailability of those features at the current time of the measurements.

Long term evolution (LTE) UE categories are characterized by The 3rd Generation Partnership Project (3GPP) releases where different UE categories support for different features and exhibit different performance 3GPP. The UE category defines a combined uplink and downlink capability. LTE release 8 defines five LTE UE categories including Category-2, Category-3 and Category-4, 3GPP Release 10 (named as Long Term Evolution Advanced (LTE-A)) introduces three new categories including Category-6 and Category-7 and 3GPP Release-11 adds four additional categories including Category-9 and Category-10 which are experimented in this paper. In this paper, four main KPIs...
(or metrics) are tracked for analysis: Download (DL) speed, latency, jitter and Packet Loss (PL) which are detailed next. These metrics are collected from UEs that are measuring their network performance, hence the crowdsourced dataset is built. The brief descriptions of the KPIs are as follows:

- **DL speed**: this KPI represents the amount of bits sent from the server to the UE in seconds.
- **Latency**: this KPI represents the time it takes for the application server to reply to a request from UE. It represents the round-trip time that is measured in milliseconds. For applications that require interactive responses (e.g. in video game applications), latency should be minimized.
- **Jitter**: this KPI is a measure of variability of latency in time. It is especially important in streaming and gaming applications where high jitter can yield interruptions or buffering of the stream.
- **PL**: this KPI describes the percentage of packets lost in comparison with packets sent. This can be due to many reasons in wireless communications and networking applications where in most cases it is due to poor signal quality.

### 2.2 Confidence interval for large-scale observations

To determine the CI of OS comparisons using the collected test speed observations, comparisons of OSs based on those collected unpaired KPI observations are calculated [39]. First, the sample means are computed as,

\[
\bar{x}(m, l) = \frac{1}{n(m, l)} \sum_{i=1}^{n(m, l)} x_{ki}(m, l),
\]

\[\forall m \in M, \forall l \in L, \forall k \in K\]  

(1)

and sample standard deviations as

\[
s_k(m, l) = \sqrt{\frac{\sum_{i=1}^{n(m, l)} (x_{ki}(m, l) - \bar{x}_{ki}(m, l))^2}{n_k(m, l) - 1}},
\]

\[\forall m \in M, \forall l \in L, \forall k \in K,\]  

(2)

where \(x_{ki}\) corresponds to KPI value of single observation \(i\) (e.g. throughput, latency, jitter values) and \(n_k(m, l)\) corresponds to total number of observations for OS-\(k\), MNO-\(m\) and UE category-\(l\). To compare OS-\(k_1\) ∈ \(K\) and OS-\(k_2\) ∈ \(K\) using unpaired observations, the mean difference \(d_{k_1,k_2}(m, l)\) is calculated as,

\[
d_{k_1,k_2}(m, l) = \bar{x}_{k_1}(m, l) - \bar{x}_{k_2}(m, l),
\]

\[\forall k_1, k_2 \in K, k_1 \neq k_2, \forall m \in M, \forall l \in L,\]  

(3)

as well as the standard deviations of the mean difference as

\[
s_{k_1,k_2}(m, l) = \sqrt{\frac{s_{k_1}^2(m, l)}{n_{k_1}(m, l)} + \frac{s_{k_2}^2(m, l)}{n_{k_2}(m, l)}},
\]

\[\forall k_1, k_2 \in K, k_1 \neq k_2, \forall m \in M.\]  

(4)

Then, the CI for the mean difference KPIs of OS-\(k_1\) ∈ \(K\) and OS-\(k_2\) ∈ \(K\) can be computed as

\[
CI_{k_1,k_2}(m, l) = d_{k_1,k_2}(m, l) \pm z_{1-\frac{\alpha}{2}}s_{k_1,k_2}(m, l),
\]

\[\forall k_1, k_2 \in K, k_1 \neq k_2, \forall m \in M, \forall l \in L,\]  

(5)

where \(\alpha\) denotes the significance level and \((1 - \alpha)100\) is confidence level, \(z_{1-\frac{\alpha}{2}}\) is the \((1 - \alpha/2)\)-quantile of \(\mathcal{N}(0, 1)\) (normal distribution with zero mean and unit variance).

### 2.3 Confidence interval for the proportions

CI for proportions method is used to calculate the PL CI values [39]. For proportions, the average probability of PL over the observation duration for all MNO-\(m\) ∈ \(M\), UEOS-\(k\) ∈ \(K\),

\[
CI_{k}(m, l) = \frac{\hat{p}_{k}(m, l) \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_{k}(m, l)\hat{q}_{k}(m, l)}{n_k(m, l)}}}{\sum_{k\in K} n_k(m, l)},
\]

\[\forall m \in M, \forall l \in L,\]  

(6)

where \(\hat{p}_{k}(m, l)\) denotes the average probability of PL of OS-\(k\) over all observations and \(\hat{q}_{k}(m, l) = 1 - \hat{p}_{k}(m, l)\) is the average probability of non-PL.

---

**Fig. 1** Number of experiments heatmap over different MNOs and UE categories in 4G networks
error is calculated as a given MNO-m ∈ distribution with Nk observation for OS-k observations’ means is of each other, the standard error for a difference between two performance test observations from different OSs are independent UE category l ∈ L are calculated. When two network performance test observations from different OSs are independent of each other, the standard error for a difference between two observations’ means is

\[
s_{k_1,k_2}^L(m, l) = \sqrt{(s_{k_1}^L(m, l))^2 + (s_{k_2}^L(m, l))^2},
\]

∀k1, k2 ∈ K, k1 ≠ k2, ∀m ∈ M, ∀l ∈ L, (6)

where \( s_{k_1}^L(m, l) \) and \( s_{k_2}^L(m, l) \) are the standard error of the observation for OS-k1 ∈ K and OS-k2 ∈ K respectively for a given MNO-m ∈ M and UE category l ∈ L. The standard error is calculated as

\[
s_{k_1}^L(m, l) = \sqrt{\frac{p_{k_1}(m, l)(1 - p_{k_1}(m, l))}{N_{k_1}(m, l)}},
\]

∀m ∈ M, k1 ∈ K, ∀l ∈ L. (7)

where the PL probability for \( N_{k_1}(m, l) \) observations is \( p_{k_1}(m, l) \). If there exists normal approximation of binomial distribution with \( N_{k_1}(m, l) p_{k_1}(m, l) \geq 10 \), the confidence interval for the PL probability can be calculated as,

\[
CI_{k_1}(m, l) = p_{k_1}(m, l)
\pm z_{1-\alpha/2} \sqrt{\frac{p_{k_1}(m, l)(1 - p_{k_1}(m, l))}{N_{k_1}(m, l)}},
\]

∀m ∈ M, k1 ∈ K, ∀l ∈ L. (8)

Therefore, the difference of two PL probabilities of OS-k1 and OS-k2 with (1-α)100% CI which is denoted as \( CI_{k_1,k_2}(m, l) \) can be calculated as

\[
CI_{k_1,k_2}(m, l) = \tilde{p}_{k_1}(m, l) - \tilde{p}_{k_2}(m, l)
\pm \frac{z_{1-\alpha/2} s_{k_1,k_2}^L(m, l)}{\sqrt{\frac{N_{k_1}(m, l)}{N_{k_2}(m, l)}}},
\]

∀k1, k2 ∈ K, k1 ≠ k2, ∀m ∈ M, ∀l ∈ L, (9)

where \( \tilde{p}_{k_1}(m, l) \) is the average PL probability over all observations. Note that based on the CI values calculated using (9) [or (5)], the statistical mean PL (or throughput, latency, jitter) difference values between OSs are significant if the interval does not contain zero with (1-α) 100% confidence.

3 Performance comparison results

3.1 Dataset and coefficient of variation comparisons

For comparative analysis of mobile OSs, extended duration version of the statistical proprietary dataset that are used in previous works [36,37] are utilized. Table 2 shows the summary statistics about the nationwide network performance test data collected in Turkey focusing on the statistics of OSs-level details. The range of crowdsourced dataset is from 01 January 2016 to 31 January 2018 that covers a period of 25 months. Unless otherwise stated, 4G cellular technology measurements are used in all the analysis below. Figure 1 shows the heatmap of the number of experiments distributed over each MNOs and UE categories. This figure shows that there is a clear imbalance between the number of observations and phone categories over all OSs and MNOs. The highest number of observations has been with the MNO-1’s Android cat-4 phones (with 369,545 observations) whereas the lowest number of observations has been with MNO-2 IOS cat-2 phones (with only 11 observations). Note also that for cat 9-10 phone categories in windows OS no observations are made for all MNOs. Figure 1 clearly indicates that in the phone category of cat-4 and in the OS of Android, higher number of network test speed experiments are observed.

Figure 2 shows coefficient of variation (CoV) values of total number of 4G measurements over different UE categories, MNOs and UE OSs for DL Speed, latency, jitter and PL KPIs in heatmap representation. CoV yields a stan-
standardized measure of dispersion for a given probability or frequency distribution. CoV value for a given MNO-\( m \in \mathcal{M} \), UE OS-\( k \in \mathcal{K} \), UE category \( l \in \mathcal{L} \) is calculated as

\[
CoV_k(m, l) = \frac{\sigma_k(m, l)}{\mu_k(m, l)}.
\]  

where \( \sigma_k(m, l) \) and \( \mu_k(m, l) \) are the standard deviation and mean over \( N_{k1}(m, l) \) observations for a given MNO-\( m \in \mathcal{M} \), UE OS-\( k \in \mathcal{K} \), UE category \( l \in \mathcal{L} \). Hence, CoV measures how spread is the dataset and is used to compare dispersion between two variables having different number of ranges (e.g. DL speed and jitter in our case). For example, if a dataset has larger deviation relative to its own mean more variability in the dataset will occur. Figure 2 shows that variations in DL speed values are much lower than variations in latency, jitter and PL. Moreover, it can generally be observed that for jitter values, there is high CoV and more spread out in comparison with CoV values of other KPIs (i.e. DL speed, packet loss and latency). Note that zero values in Fig. 2 correspond to non-availability of the data in the considered cell. In the remainder of the paper, for comparisons of OSs, CIs for IOS-Android and Windows-Android mean KPI values are calculated. Note also that CI values for UE categories such as Cat-1 and Cat-0 are not shown due to lack of available network performance test data for corresponding OSs. For example, Cat-1 lacks IOS OS test experiments and Cat-0 lacks both Android and Windows OS experiments.

### 3.2 Effect of UE category and OS on KPI of MNOs

In this subsection, the effect of UE category (ranging form Cat-2 to Cat 9-10, while excluding Cat-0, Cat-1 due to non-existence of paired observations) and UE OS (Android, IOS and Windows) on the KPI for all three MNOs using 4G cellular networks is investigated. In 3GPP standards, there are 12 defined LTE UE categories where each UE categories have wide range of supported parameters and performances [38]. For example, LTE UE Cat-1 devices are not supporting Multiple InputMultiple Output (MIMO) whereas Cat-5 and above categories can support \( 4 \times 4 \) and above MIMO. In order to observe the difference of the effect of OS on different
Android vs. iOS: a comparative analysis over mobile operator infrastructures based on...

Fig. 4  a MNO-1 latency v.s UE categories, b MNO-2 latency v.s UE categories, c MNO-3 latency v.s UE categories d MNO-1 CI mean latency difference v.s UE categories, e MNO-2 CI mean latency difference v.s UE categories, f MNO-3 CI mean latency difference v.s UE categories

UE categories for each MNO, CI values that are calculated using (5) (for DL speed, latency, jitter values) and (9) (for PL values) are also plotted.

DL speed CI: Figure 3 shows the DL speed boxplots and corresponding CI values for DL speed comparisons of iOS, Android and Windows OS phones over UE categories. For MNO-1, from Fig. 3a, d, it can be observed that IOS OS is better than Android OS for Cat-4 and Cat 6-7 devices. For Cat-2 and Cat-3 UE categories, IOS is worse than Android. On the other hand, Windows OS performs worse than Android in all UE categories. For MNO-2 of Fig. 3b, e, it can be observed that IOS performs better than Android for Cat-4, whereas for Cat-2 and Cat 6-7 devices it is vice-versa and no conclusions between Android and IOS can be drawn for Cat-3 devices. When Android and Windows OSs are compared, again lower DL speed values for Windows phones in all UE categories are observed. For MNO-3 of Fig. 3c, f, it is observed that IOS is better than Android for Cat-3 and Cat-4 devices, whereas Android is better for Cat 6-7 devices and no definitive decisions can be given for OSs comparisons over Cat-2 devices with 90% confidence level. Similarly, Android performs better than Windows phones in all UE categories for MNO-3 as well.

Latency CI: Figure 4 plots the latency boxplot values and CI for mean latency difference values over UE categories for comparisons of different UE OSs over UE categories. For MNO-1 latency values, it can be observed from Fig. 4a, d that for Cat-2 and Cat 6-7 devices, one cannot decide with 90% confidence level whether IOS or Android OS is better due to inclusion of zero in CI. Note that Cat-2 for latency values of MNO-2 has low number of experiments (around 11 as given in Fig. 1). Therefore, large values in both standard deviation and mean are observed due to non-availability of large-scale experiments on that set. However, for Cat-4 devices (which is the most common category type in the observed dataset), IOS OS performs better than Android for MNO-1 in terms of latency which cannot be directly inferred from Fig. 4a, but can be distinguished clearly in Fig. 4d. On the other hand, for Cat-3 devices Android is better than IOS. At the same time, for all UE categories, Windows OS has worse latency values than Android OS.
For MNO-2 latency comparisons of Fig. 4b, e, it is again observed that Windows OS phones perform the worst, whereas IOS OS phones perform the best compared to Android in Cat-4 devices. On the other hand, Android is better than IOS in Cat-2 devices. No conclusions can be drawn between IOS and Android for Cat-3 and Cat 6-7 devices since measurements give insufficient information due to inclusion of zero value in CI. Similar trends for Windows OS devices can also be observed for MNO-3 in Fig. 4c, f where Windows OS phones have the largest latency values. When IOS and Android are compared, Cat-2 and Cat-3 devices Android OS performs better than IOS, whereas for Cat-4 and Cat 6-7 devices, IOS is better than Android.

**Jitter CI:** Figure 5 shows jitter boxplot values and CI for mean jitter difference values over UE categories when different OSs are compared. Fig. 5a, d show the jitter values for MNO-1 where Cat-3, Cat-4, Cat 6–7 and Cat 9–10 UE categories in IOS devices have higher jitter values. For Cat-2 devices, one cannot conclude with 90% confidence whether Android or IOS devices are better than one another. Note also that similar to latency observations, Cat-2 for jitter values of MNO-2 has low number of experiments (around 11 as given in Fig. 1). Therefore, large values in both standard deviation and mean are observed due to non-availability of large-scale experiments on that set. For MNO-2 from Fig. 5b, e, it can be observed that for Cat-3, Cat-4, Cat 6–7 and Cat 9–10 UE categories, IOS devices have lower jitter than Android devices and for Cat-2 devices one cannot make a comparative decision. For MNO-3, from Fig. 5c, f, it can also be observed that for Cat-3 and Cat-4 devices, IOS phones are better than Android. On the other hand for Cat-2, Cat 6–7 and Cat 9–10 devices, one cannot make a conclusion on performance comparisons between Android and IOS devices.

**PL CI:** Figure 6a–c show the CI values for mean PL difference of UE categories for comparisons of different OSs (i.e. Android, IOS). In general, for all MNOs, Android UEs have higher PL ratios compared to IOS in Cat-3 and Cat 9-10 devices whereas for Cat-2 UEs in MNO-1 and MNO-3, Cat-4 UEs in MNO-2 and Cat 6-7 UEs in MNO-3, one cannot conclude with 90% confidence that Android or IOS OSs is better than one another. For MNO-1 in Fig. 6a, it can be observed that IOS OS UEs have higher PL ratios compared to Android OSs (i.e. performs worse) in Cat-4 and Cat 6–7 devices and for MNO-2 in Fig. 6b, IOS is better than Android in Cat-2 devices.

**Main takeaways:** In general, when considering the effects of UE categories on MNO performances, it is observed that Windows OS UEs perform poorer than Android OS UEs for all MNOs and UE categories during the observation period. On the other hand, depending on the UE category, the performances on MNO-1, MNO-2 and MNO-3 may vary when IOS and Android OSs are compared for different KPIs. The PL represents the percentage of packets lost in comparison with transmitted packets and is mostly a network performance parameter. Therefore, there may be different reasons for observing PL such as Radio Access Technology (RAT) changes, temporary loss of service, connection termination or Radio Resource Control (RRC) state transitions, diurnal patterns just to name a few as also stated in [21] and [40]. In addition to these possible factors that are observed previously in the literature, the analysis results in this paper have also indicated the clear differences in PL performances between UE categories even though PL is commonly considered as a metric due to problems in network side. In addition to previously described reasons in the literature as described before, some of the additional reasons for this discrepancy can be due to receive buffer size differences between UE categories or correct delivery of packets from the network but execution of higher drop rates at UE side. Using the whole dataset for all MNOs, a summary of performance characteristic of the different UE categories is given in Table 3 where “N/D” represents no decision.

### 3.3 MNO and OS Comparisons over different KPIs

Figure 7 shows DL speed, latency and jitter values in both boxplot and corresponding CI values plot. Note that when comparing technologies, first the dataset is separated based on groups of similar MNOs (i.e. MNO-1 and MNO-2) which ensures that all portions of the dataset (e.g. including all categories) are used for CI calculations. Later, the means of OSs per technologies are compared using CI for large-scale observations and for the proportions methodologies.

For DL Speed, as can be observed from Fig. 7a, d, Android performs better than IOS UEs and when Windows and Android are compared, it can be observed that Windows OS UEs have worse DL speed rates for all MNOs. On the other hand, from latency figures of Fig. 7b, e and jitter figures of Fig. 7c, f, different results for Android and IOS are observed. Android UEs have higher latency compared to IOS UEs for MNO-1 and MNO-2 (no conclusions can be drawn for MNO-3) and Windows have worse latency performance results than Android over all MNOs. When jitter values are compared as in Fig. 7c, f, IOS performs better than Android for MNO-2 and MNO-3 but is worse than Android in MNO-1 (note that jitter values in Windows OS UEs are not observed due to absence of this experiment data over those phones). When PL performances over all MNOs are compared as in Fig. 7g, it can be observed that for MNO-2 and MNO-3, IOS performs better than Android, whereas for MNO-1, Android is better than IOS UEs. This observation is similar to MNO comparisons in terms of jitter values.

**Main takeaways:** A summary of the best OSs over all considered MNOs and KPIs is given in Table 4 using the information in Fig. 7. In this table “N/D” again represents no decision. From this table, when considering the effects of
Fig. 5  a MNO-1 jitter v.s UE categories, b MNO-2 jitter v.s UE categories, c MNO-3 jitter v.s UE categories, d MNO-1 CI mean jitter difference v.s UE categories, e MNO-2 CI mean jitter difference v.s UE categories, f MNO-3 CI mean jitter difference v.s UE categories

Fig. 6  CI mean PL difference v.s UE categories a MNO-1, b MNO-2, c MNO-3
MNOs and UE’s OS, it can be observed that Android UEs have better DL speed performances and worse latency and jitter performances compared to IOS. This signifies that for applications requiring low latency and jitter (e.g. in streaming and video applications), IOS phones have major advantages for all MNOs in Turkey. However, for high speed requiring applications (e.g. video downloads) Android phones have higher performance compared to others. At the same time, Android OSs have greater KPI advantages compared to Windows OSs in all MNOs as observed in Fig. 7. In terms of PL and jitter values, IOS devices have performed better in two operators, namely in MNO-2 and MNO-3 but has performed poor in MNO-1. These results again signify that when UEs are willing to obtain MNO services that require low jitter and low PL, not only the choice of OS of phones is important but also the choice of MNO can be significant in some scenarios.

3.4 Visualization of the nationwide analysis results

In this subsection, CI values will be visualized to facilitate efficient data exploration at a geographically large scale of the analyzed results for 4G cellular networks in Turkey. Fig. 8 shows the CIs for mean DL speed difference of IOS and Android for 4G networks in all cities of Turkey. In this figure, if IOS performs better than Android in a city, the city is marked as blue, if Android performs better than IOS, the city is marked as red and if CI of mean difference includes zero, the city is marked as white.

Similarly, Fig. 9 visualizes the mean latency performance difference comparisons of IOS and Android using Folium and leaflet map visualization in Turkey. As can be observed from this figure, IOS performs the best in red marked cities and Android is the best in blue marked cities whereas no conclusions can be drawn in white marked cities in terms latency performance.

Main takeaways: The results from Figs. 8 and 9 reveal that depending on the region of operation, there can be significant differences between OSs’ performance (e.g. IOS is better in majority regions of the country compared to Android in terms of DL speed whereas Android and IOS better cities are equally distributed throughout the country in terms of latency values). It is also interesting to observe that Android performs the best when comparisons are grouped according to each MNOs as previously shown in Table 4, whereas when comparisons are grouped according to cities, IOS performs mostly better than Android in most of the cities and in larger geographical areas in Turkey. In this respect, the performances of OSs depend highly on which grouping conditions (whether it is based on city, MNO, phone category, etc.) are utilized during comparison operations. In summary, using the insight drawn from the studied statistical methodology outlined in this paper, network equipment vendors, device manufacturers, OS and application developers, MNOs and users in the whole mobile market ecosystem can distinguish their strengths and weaknesses in certain regions of the country, phone categories.

3.5 Business impact and consequences of OS comparisons

In this section, the consequences of the studied comparative analytic approaches are studied in terms of user experience, business benefits and investment plans for various actors in the mobile industry ecosystem, namely device manufacturers, OS developers, MNOs, network equipment vendors and users. Note that the development of a new OS software and necessary critical updates continuously done in cooperation with device manufacturers are time consuming processes and can be costly if appropriate measurements and experiments are not taken into account. Therefore, OS developers and device manufacturers need to carefully and intelligently perform business planning to support the necessary investments while taking into account their competitors strengths and weaknesses. For example, the economical impact of comparing DL speed, latency, jitter and packet loss over various phone categories and MNOs can change the marketing plan of OS developers and device manufacturers. This can help them to enhance business decisions in a fine-grained scale, e.g. OS developers and device manufacturers may invest in enhancing the capabilities of mobile phones that cannot well match with their software characteristics in terms of performances or cooperate with MNOs to promote devices and OSs that are well aligned with their performance metric outcomes.

The analytic methods and obtained results done in this paper can also inform OS developers and device manufactur-
Fig. 7 UE OSs comparisons for different MNOs 

(a) DL Speed Boxplot vs MNOs, (b) Latency BoxPlot vs MNOs, (c) Jitter BoxPlot vs MNOs, (d) CI mean DL Speed difference vs MNOs, (e) CI mean latency difference vs MNOs, (f) CI mean Jitter difference vs MNOs, (g) CI mean PL difference vs MNOs.

ers to invest in new products/services and planning strategies. For example in our analysis, in terms of DL speed values, IOS OSs perform worse than Android OSs in cat-2 and cat-4 devices as given in Table 4. For this reason, development units of IOS OS and device manufacturers can concentrate on improving their OS capabilities in cat-2 and cat-4 capable devices for throughput-critical services such as video on demand (VoD) or streaming services. Similarly, Android OS developers can invest in enhancing their online gaming experience together with their device manufacturers due to their inferior latency values in cat-4 and cat 6-7 phone categories as observed in Table 4. This can also give an intuition about how
Table 4 A summary of better OSs grouped over different MNOs and KPIs

|                      | MNO-1     | MNO-2     | MNO-3     |
|----------------------|-----------|-----------|-----------|
| DL speed             | Android   | Android   | Android   |
| Latency              | IOS       | IOS       | N/D       |
| Jitter               | Android   | IOS       | IOS       |
| PL                   | Android   | IOS       | IOS       |

Fig. 8 Android and IOS comparisons in terms of DL speed performances among all cities in Turkey. [Figure is best viewed in colors]

Fig. 9 Android and IOS comparisons in terms of latency performances among all cities in Turkey. [Figure is best viewed in colors]

the business strategies of OS developers and device manufacturers may be impacted when evolution from matured phone categories to cutting-edge phone categories occur.

Another example of analysis results grouped by each MNOs reveal the following conclusion: Due to clear disadvantage of IOS OS in DL speed compared to other Android OS as given in Table 4, IOS OS developers and device manufacturers need to investigate the reasons behind this fall-back. One strategy can be to focus their efforts on extensive experiments and evaluation tests focusing on DL speed improvements. During our investigation of how phone categories can affect OSs performance differences, new discoveries are also detected. For the more basic phone category of cat-2, the observed DL speed and latency performances are better in Android OS, whereas for the more advanced phone category of cat 6–7, IOS OSs have performed better. This signifies that in terms of DL speed and latency performances, IOS OS developers and device manufacturers have advanced their performances much better than their Android competitors as new phone categories are introduced jointly.

The location itself can also affect the enhancement strategies of OS developers and device manufacturers. For example, for DL speed values, IOS is better than Android in most of the regions of Turkey as shown in Fig. 8. This awareness paves the way for investing and enhancing the capabilities in selected cities/regions of the country in a strategic manner. Nonetheless, the performance advantage of a single OS over one KPI (e.g. latency advantage of IOS OS in southern city of Antalya) does not guarantee that same OS will outperform others in remaining KPIs (e.g. DL speed advantage of Android in Antalya). New services that will be introduced in certain regions will mostly likely depend on how well the OS performs in that region.

With the observation of potential performance fall-backs in different geographic locations, phone categories and MNOs compared to their competitors, OS developers and device manufacturers can jointly improve their KPI performance. This will also result in higher user quality-of-service (QoS) and quality-of-experience (QoE). This can benefit end-users who are eager to obtain OS and devices that yield higher performance inside their regions of the country. Long-term accurate and persistent comparisons allow OS developers, device manufacturers, network equipment vendors and MNOs to anticipate future Capital expenditure (CapEX). This enables them to make more confident strategic decisions for planning investments while increasing business efficiency.

In summary, our analysis results reveal two major suggestions for OS developers, device manufacturers, MNOs and network equipment vendors: First, together with awareness of OS capabilities in different MNOs (or network infrastructures) and phone categories, they can together improve their services (e.g. online gaming, video on demand or streaming music services) with minimum OS software and hardware investments by considering their major strengths and weaknesses under different KPIs in comparison to their competitors. Second, based on location performance differ-
ferences, some of the mobile players in the ecosystem can join forces and cooperate in certain regions of the country by marketing more advantageous phone categories and OSs to boost their profits and reduce redundant investments. This can also help to attract more consumers for all the mobile actors of the telecommunication ecosystem.

3.6 Extended data sources for enhanced comparative analysis

Note that the proposed statistical analysis method in this paper provides a new way of OS comparisons based on various factors. In general, the analysis done in this paper can be extended by utilizing enhanced dataset that can also accommodate more diverse metrics such as channel conditions at the time of the experiments, utilized firmware details, connected eNodeBs, etc. by joining data from data warehouses of different mobile ecosystem actors (MNOs, vendors, device manufacturers, OS developers, etc.). In general, the utilized crowd-sourced data in this paper can be enriched by combining the available data from three separate systems of a general telecommunication infrastructure: Information Technology (IT), network and application systems.

− **IT systems data** can be collected from various IT systems such as Customer Relationship Management (CRM) systems, billing systems or customer care services. Some examples are basic information on customer, account, and user profile and characteristics data, billing data, business consumption data, social media and promotions/campaigns data from marketing/sales department, etc.

− **Network systems data** (from both wired and wireless networks) include network equipment data, Call Detail Records (CDRs), eXtended Data Records (XDRs), Machine-to-Machine (M2M) data, traffic data (both signalling and payload data), Operations Support Systems (OSS) data (network events (failures, alarms), network performance data, etc), voice, SMS, networking service data, UE mobility and location updates, QoS parameters data collected from access, transport and core networks (e.g. from Access and Mobility Management Function (AMF), NG–RAN, Authentication, Authorization and Accounting (AAA), Home Subscriber Server (HSS) servers, etc.).

− **In applications/services data**, data is from products and services (e.g. online mobile payment, online music and e-wallet applications or vehicle tracking, power grid information and healthcare services, other value added services, etc) provided by the telecommunication operators that include user data (e.g. user access modes, addresses, timestamps, business preferences, consumption habits, customer care agent’s data). Note that the underlying structures of these data can be in complex data format (either unstructured (text, images, videos), structured or semi-structured).

4 Conclusions and future work

In this paper, end-to-end network performance comparisons of OSs over different UE categories and MNOs are studied. The results are compared using the previously collected dataset of test speed that ranges 25 months of duration in Turkey. The analysis results give insights into end-to-end network performance of OSs for each MNO and comprehensive understanding of the statistical KPI differences (using CI evaluations over DL speed, latency, jitter and PL values) of OSs based on UE categories using long duration and nationwide real data measurements. The analysis results indicate that Android devices perform better in terms of DL speed values among all MNOs, whereas IOS devices are better in latency. On the other hand depending on the UE category, the performances of IOS and Android may vary based on different MNOs and KPIs. Note that as a future work a controlled environment can be created to have the eNodeB and the wireless communications context identical (within some acceptable error) for all the measurements. This can ensure to experimentally pinpoint the role of the UE OS or category, in the overall performance (DL speed, latency, jitter, etc) and help in comparisons with the previously collected crowd-sourced measurements. Other observations and parameters such as RF characteristics of the user environment, firmware version of the UEs, and the RF chip-set between different OSs, etc. that can be obtained from a larger data sources to connect as described in Sect. 3.6 can also be included in the calculations of the CI values. Finally, a similar analysis can also be run for further cellular technology comparisons using 5G networks when they are commercially available large-scale in Turkey.

Acknowledgements This work was partially funded by Spanish MINECO grant TEC2017-88373-R (5G-REFINE) and by Generalitat de Catalunya Grant 2017 SGR 1195.

Declaration

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Ahmed, E., Yaqoob, I., Hashem, I. A. T., Shuja, J., Imran, M., Guizani, N., et al. (2018). Recent advances and challenges in mobile big data. *IEEE Communications Magazine*, 56(2), 102–108.
36. Yildirim, A., Zeydan, E., & Yigit, I. O. (2020). A statistical comparative performance analysis of mobile network operators. *Wireless Networks*, 26, 1105–1124.

37. Zeydan, E., & Yildirim, A. (2020). Reliability comparisons of mobile network operators: An experimental case study from a crowdsourced dataset. *Turkish Journal of Electrical Engineering & Computer Sciences*, 28, 380–393.

38. 3rd Generation Partnership Project (3GPP). “User Equipment (UE) radio access capabilities” In 3GPP TS 36.306 V15.6.0 (2019-09), from [https://www.3gpp.org/ftp/tsg_ranWG2_RL2Specifications/201909_final_specs_after_RAN_85/36306-f60.docx](https://www.3gpp.org/ftp/tsg_ranWG2_RL2Specifications/201909_final_specs_after_RAN_85/36306-f60.docx).

39. Jain, R. (1990). The art of computer systems performance analysis: Techniques for experimental design, measurement, simulation, and modeling. Wiley.

40. Baltrūnas, D., Elmokashfi, A., & Kvalbein, A. (2015). Dissecting packet loss in mobile broadband networks from the edge. In 2015 *IEEE conference on computer communications (INFOCOM)* (pp. 388–396).

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Engin Zeydan received his Ph.D. degree in February 2011 from the Department of Electrical and Computer Engineering at Stevens Institute of Technology, Hoboken, NJ, USA. Previously, he received his M.Sc. and B.Sc. degrees from the Department of Electrical and Electronics Engineering at Middle East Technical University, Ankara, Turkey, in 2006 and 2004, respectively. Dr. Zeydan has worked as an R&D engineer for Avea, a mobile operator in Turkey, between 2011 and 2016. He was with Turk Telekom Labs working as a Senior R&D Engineer between 2016 and 2018. He was also part-time instructor at Electrical and Electronics Engineering department of Ozyegin University between 2015 to 2018. He is currently with Communication Networks Division of the Centre Tecnologic de Telecomunicacions de Catalunya (CTTC) working as a Senior Researcher. He received Best Paper Award from the Network of Future Conference in 2017. His research interests are in the areas of telecommunications and big data networking.