Evolution of Measurement of the Single Piece Mail Transit Time

Ryszard Kobus[1] and Frank Raudszus[2]

[1] *National Institute of Telecommunications, Warsaw, Poland*
[2] *Bundesnetzagentur, Mainz, Germany*

Abstract—The article presents the evolution of measurement of the transit time of priority mail methodology used for postal services quality assessment from the point of view of home and small business senders based on European postal standards. The paper also considers the possibility of test cost reduction by further changes in the measurements methodology.

Keywords—European standards, measurements of transit time, postal services.

1. Introduction

Delivery of single piece mail is a low-cost service addressed to private persons or small entrepreneurs enabling them to send paper documents or items. It is usually posted by letterboxes or postal offices and its route is not registered. Therefore, the service is not subject to complaint and if an item is lost or its delivery is delayed the client cannot receive compensation.

For many years, the quality of service has been monitored by measuring the transit time of test items. The European UNEX measurement system has been operating since 1994 and it tests the transit time of the cross-border priority mail.

2. The First European Standard

The European Commission emphasized the need to establish common rules for the development of community postal services and the improvement of quality of service (QoS) [1]. The Commission identified that the QoS measurement system should include:

- independent end-to-end measurement capabilities,

- a focus on cross-border delivery service performance,

- a single, uniform and reliable system for monitoring delivery service performance within the European Union.

The Commission acknowledged that the many postal traditions and different cultures in Europe would not allow the establishment of one common unified measurement system for domestic mail. Therefore, the national postal systems should have sufficient degrees of freedom to reflect local needs and peculiarities. On the other hand, they should fulfill a defined set of minimum requirements to satisfy the information interests of the national regulatory authority, postal customers as well as postal operators.

The first European standard defining the measurement system was drafted by technical committee CEN-TC331 and published in 2002. The EN 13850 [2] defines the measurement method of the transit time of end-to-end services for single piece priority mail and first-class mail for domestic and cross-border services. The standard was addressed to EU members, i.e. to countries with big postal flows. Standards [3] for second-class mail (EN14508) and for bulk mail (EN14534) measurements were published with a slight delay.

The transit time tests were based on measurements of letters posted and received by test panel, recruited and managed by independent research organization. The mailing should be distributed in at list 30 contractual postal areas. The standard does not define directly the size of the test sample and the size of the test panel but states that they should ensure the accuracy of 1% for domestic and 5% for cross-border measurements.

To ensure high accuracy of the test sample, it should be representative of the real mail. This is achieved by geographical stratification based on the following categories:

- urban – cities including their outskirts,

- rural – smaller cities and communes,

- local I – mail sent to and delivered within the same urban city,

- local II – mail sent within and between rural parts of the same catchment area,

- distance I – within a straight-line distance of 200 km,

- distance II – above a straight-line distance of 200 km.

The basic definitions of outskirts, urban, rural, local and distances are described in the EN standard. The domestic measurements may take national peculiarities into account, by agreement with the regulatory authority. By default, as the city is assumed to be a town above 50,000 inhabitants but in Poland it has been assumed that urban are cities above 10,000 inhabitants.

The sample of test items shall be representative of the real mail for at least the 10 geographical parameters based on: point of posting and point of delivery, for example:

- urban – urban – distance I,

- urban – urban – distance II,

- urban – rural – distance I,

- urban – rural – distance II, etc.

The test items shall have correctly written addresses in according to recommended template and with respect to addressee, delivery address with postcode and for cross-border mail the destination country. The test sample shall represent a statistical sample of the real mail for all discriminant mail characteristics. Generally, the following discriminant mail characteristics should be assumed as a minimum and shall be estimated through real mail studies:

- day of induction as day of the week,

- time of posting for urban areas (it is only important if more than 1 collection time is published),

- all typical methods of delivery.

In real conditions, some test templates can be considered negligible especially in domestic mail. The service delivery at P.O. boxes is often omitted. In Poland, the test mail is paid only by stamps, because other payment methods are rare.

The implementation of EN 13850 was difficult, hence the TC331 workgroup published an implementation guide which describes all stages of preparing and starting the measurements in detail. It should be noted that the application of the rule has limitations, for example:

- the EN 13850 European standard may not be suitable for the measuring of very small volumes of mail;

- a test flow cannot noticeably affect the overall mail traffic. Usually it is assumed that test flow cannot increase the total flow by more than 0.2%;

- the standard may not be suitable for operators with limited coverage. Therefore, the covered geographic area should be large enough so that it can be divided into 30 contractual postal areas;

- this standard is not applicable for measuring the end-to-end transit time distribution of large bulk mailers' services and hybrid mail, which require different measurement systems and methodologies.

The standard makes it possible to conduct measurements with or without electronic system for item tracking, and thus to validate results. In Europe, the transit time of cross-border letters is measured in the UNEX system [4], [5] with Radio Frequency Identification (RFID) active tags. The gates for events registration are installed in sorting centers in all Europe. Such a route tracking system can be used to measure transit time for each part, but this functionality is not covered by this standard.

Generally, the standard is used to measure the achievements of the postal operator in the form of two annual indicators of quality of service.

2.1. On-time Performance

The on-time performance is described as the percentage of postal items delivered within the defined service standard expressed in days. A report should present the level of on-time performance accuracy achieved in the test period.

2.2. Cumulative Distribution of Delivery Days

The cumulative distribution of delivery days is a percentage factor of mail delivered within a given period, from 1 to 10 days. All postal items delivered up to 30 days shall also be considered in the calculations.

It should be added that the resulting indicators cannot be used for direct comparison of achievements by other operators due to different conditions, i.e. geographic areas handled or postal flows, which can make the results incomparable.

3. EU Countries with Less Letter Flows

In 2004, a group of 10 new countries joined the EU. In this group, there are also countries with relatively small populations and thus with low mail flows. An additional problem is that the priority mail service does not enjoy great popularity. It requires reducing the random sample test mail. The developed A1 supplement [6] temporarily solved this problem, but in the meantime work was undertaken to develop a new edition of the standard.

The A1 proposes two methods to decrease the annual test:

1. Extending the period of measurement to two or three years. This way, the annual real mail volume can be smaller while keeping full accuracy. The disadvantage of this approach is that the results for reaching full accuracy after 2 or 3 years should not be reported until these years have passed. After that, the results should continue to be reported annually on a 2- or 3-year rolling basis.

2. Case based on accuracy. The accuracy depends on the sample size which should be as big as necessary to meet the accuracy requirement for each measurement result. If the real mail volume in a given field of study is lower than a certain threshold and the on-time performance is above a certain limit, then the sample size could be decreased.

Figure 1 shows that if on-time performance is higher than 85%, then the sample size can be decreased without loss of accuracy. The expected on-time performance can be estimated based on previous results.
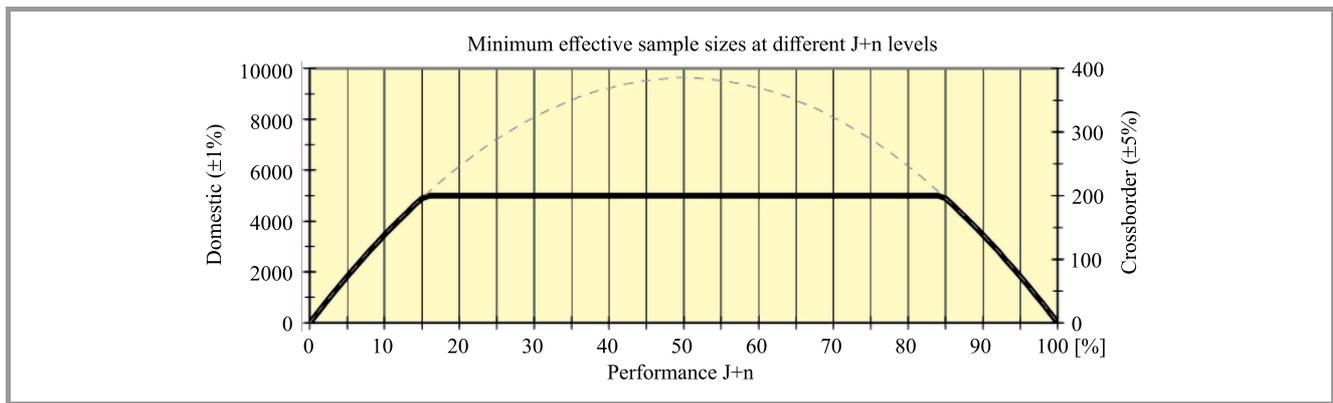
Fig. 1. Possibility of decreasing the sample size [6].

4. The Second Edition of EN 13850

In 2006, a full revision of standard EN 13850 [7] was started and huge modifications were planned. The aim was to increase the flexibility of standards through better adaptation to a random sample of actual traffic and to widen of application area.

The revision team published a survey asking about types of characteristics that are monitored in the real mail studies and their impact on measurements results, which also included questions about the type of geographical stratification used as well as how many cells were incorporated into the design. Results of this survey were implemented in the revision.

4.1. Highly Discriminant Mail Characteristics

The mailing characteristics have to be reviewed at least every three years. The list of possible criteria is given in the standard, and all of them should be checked whether they can be regarded as discriminant or not. The following list of possible characteristics that can be evaluated as a minimum:

1. Mail characteristic referring to the induction/delivery point:

   - type of geographical area by: urban, rural,

   - type of payment by: stamped, metered, postage paid,

   - type of induction by: mail street box, post office, collection from sender's premises, induction in sorting centers,

   - time of posting – only in the case of more than one collection per day,

   - type of delivery by: street address, P.O. box, delivery to receivers' premises.

2. Mail characteristics referring to the test letter itself:

   - formats by at least two modes,

   - weight steps by at least two modes,

   - addressing method by: hand written, typed,

   - weekday of induction.

One can note that the distance is no longer a discriminant mail characteristic in the presented list. Many postal systems use only big automated sorting centers. This means that the item sent locally and at a distance of approx. 200 km may travel on a similar route. Anyway, research conducted in Poland showed that for domestic mail distance has a significant impact on time performance.

4.2. Minimum Sample Size Accuracy

The revised standard gives the minimum sample size (MSS) of exactly 9,625 items, which shall be taken for a domestic measurement system. If expected performance level is greater than 50%, the minimum sample size may be reduced. For example, the 90% performance level can be taken by 3,500 items for domestic measurement system and only 1,850 items for 95% performance level.

The minimum sample size for $50-97.5\%$ performance is given in the tables as separate values for domestic and cross-border measurement systems. Generally, the minimum sample size is estimated to ensure 1% accuracy for domestic system and 5% for cross-border. All possibilities to reduce the size of the annual test given in the A1 supplement were transferred, and even extended in second edition of standard. The four categories of countries are defined according to domestic and separately cross-border mail flows. For each category, the optimal solution is proposed. The example for the domestic system is:

- Category 1 – large size mail volumes, i.e. with total annual real mail volume above 500 million mail pieces. Measurement without restriction;

- Category 2 – medium size mail volumes, i.e. with total annual real mail volume of 200–500 million mail pieces. Measurement for countries above J+n[1] performance of 85% – fixed sample size of 4,950 and for countries below J+n performance of 85% the bounded sample size is recommended;

[1]The delivery of the n-th day after posting (J).

- Category 3 – small size mail volumes, i.e. with a total annual real mail volume of 1.5–200 million mail pieces. Measurement for countries above J+n performance of 90% – fixed sample size of 3,500 and for countries below J+n performance of 90% the bounded sample size is recommended;

- Category 4 – very small size flows, i.e. for all flows with volumes below 1.5 million mail pieces per year. Measurement: in case of domestic measurement systems, the test mail can increase the real mail volume in the total field of study by more than 0.25%, undermine the neutrality of such a measurement. Broadening the field of study is recommended here, for example by including further operators.

4.3. Calculation of Accuracy

Three methods of of accuracy estimation are proposed.

Normal approximation. In most cases, the normal distribution will be an appropriate approximation of the binomial distribution. This simple normal confidence interval is symmetrical and easy to use. There are some restrictions on the use for high performance levels when the normal approximation can work poorly even with moderate sample sizes.

But for domestic measurement systems with performance levels up to 96%, the normal confidence interval can be used. In this case, usually at least 50 delayed items are registered by the measurement system. Annex A of the standard explains this problem in detail.

Agresti-Coull approximation. This estimation method is an improved method based on the normal distribution and can be used for all sample sizes with at least 40 items, which is easy to meet. This adapted normal distribution confidence interval is asymmetrical and can be used without sophisticated statistical software.

Inverse beta approximation. This improved estimation method uses the inverse beta function. It is based on the beta distribution, which is the continuous form of the binomial distribution. The inverse beta function is implemented in many software packages for data analysis. For example it is part of Microsoft Excel edition 2010 or newer. The beta distribution confidence interval is asymmetrical and easy to use.

Geographical distribution of the panel. The panels of senders and receivers are dispersed over a geographical recruitment grid, based on postal areas served by the operators. The geographical distribution of the panel is carried out according to random sampling on the whole of the geographical area defined in the field of study. In this range, the standard is very flexible and describes rules for the geographical distribution of the panel both in small countries and big ones.

For operation in bigger countries, a big panel of over 90 panelists distributed in at least in 30 areas is dedicated. In this case, there is no important change compared to the previous edition.

For small countries, a small panel (10 to 90 panelists) and from 4 to 30 postal areas are proposed. The new thing is that in this case number of postal areas is related to the number of panelists and presented in the table. The weekly workload of panelists should be limited to the level of typical user.

The standard proposes a maximum load of 12 letters per week for any domestic sender and 12 letters for any domestic receiver. But the average load of receiver is restricted to 6 letters per week during their time of participation in the measurement period. For a business address a bigger load is allowed – up to 24 letters per week. The weekly workloads given above are the maximum. In many cases, especially in countries with smaller postage flows, the actual workloads should be much lower.

4.4. Design Basic

The design of the measurement system should ensure that the test letters are allocated as a representative sample of all single piece priority mail in the field of study. The best way to achieve a representative sample would be to take as a simple random sample of real mail letters and observe their transit time. Unfortunately given the high measurement accuracy requirement such a solution is unrealistic due to technical reasons. Instead, pre-fabricated test letters are used for measurements and they are sent and received by a group of selected panelists. This design approach requires that the test letters that are added to the existing real mail stream do not differ from it in each mail characteristics, which have a significant influence on the transit time result. The characteristics and modes, which are discriminant, depend in practice on the detailed operation of the dedicated postal system operator. Therefore, a factor which is important to one operator could be not important to others.

The standard describes how to check if a characteristic is discriminant. The test is based on comparison of at least two modes of the characteristic and evaluating their impact on transit time results.

The standard is based on real mail flows and its design is determined by a system of real mail studies which estimate these flows. The real mail studies are performed before or parallel to the first test measurement period. The real mail studies make it possible to:

- consider all single piece priority flows of a given field of study,

- collect statistics on single piece priority real mail flows and real mail characteristics according to requirements of the statistical design.

The real mail studies can be implemented either by the postal operators themselves or by an external body, but they have to be independently audited.

Under the standard the distribution of real mail flows corresponding to certain mail characteristics can be estimated using existing logistic or management data available in the

postal system. These alternative solutions can decrease the cost of design and accelerate the acquisition of necessary data. They may also ensure greater accuracy of estimation than in real mail studies if data is collected by the operator on regular basis.

The standard recommends that real mail studies should be performed at least once every three years. And it is a proper period if the postal market is stable but if postal market is fast-changing then the period should be shorter.

4.5. Implementing Standard

An implementation guide, which was earlier a separate document is now implemented to the standard as Annex H. It contains a scenario which describes successive stages of implementation, from the survey planning phase through its implementation, to reporting of results and auditing. Planning and implementation phases are very time-consuming. The preparation of the test mail survey can take 9 to 18 months. Considering together the measurement period and time of data analysis as well as reporting, the first regular report may be available even after 33 months.

Detailed rules related to all stages of measurements phases and auditing process have been discussed. Furthermore, the attached examples allow for better understanding of the requirements of the standard. Generally, particular attention was paid to the problems of measurements accuracy and corrective weighing.

4.6. Measurements in Multi Operators Environment

Initially, the standard was adapted to the measurements on the liberalized postal market, which in practice means that it should be ready to conduct measurements in the environment of many operators. To fulfil this condition each panelist sender should be able to freely choose the best postal operator for posting test letters. This option was removed from the final draft, because no country except Germany was interested in such a feature.

4.7. Monitoring the Cross Border Priority Mail

The high cost of services forces one to consider the possibility of replacing test letters by the monitoring of real mail transit. This cost effective method can be based on observation of existing components or added in a typical posting technological process. Based on such an assumption, a few testing methods were selected to make a detailed investigation.

The first method is based on information retrieved from sorting machines with proper software. The method seems not expensive, except for a few countries which sort manually cross border priority mail. In addition, there are issues with the compatibility of machines of different generations. The second approach is based on reading the day of posting, in the destination postal office/sorting center. The date can be retrieved from the stamp, timestamp or franking image. Unfortunately, franking machines and digital stamps are

not popular in some countries. In addition, the day printed by franking machines or placed on a digital stamp, is the accounting day of operation, which can be different than the day of posting. Additionally, the franking machines are not used by private senders and small businesses.

This solution does not require any additional components or installation in the postal infrastructure, nor any additional technological operations in the country of origin. But in destination countries new scanners are need or the existing ones need to be modified.

The third approach is based on passive RFID tags which are hidden or placed inside the envelope. The solution needs RFID scanners in the country of origin and destination to collect data. Fortunately, the cost of RFID scanning is low, while passive RFID tags are also cheap. The big issue is the tags secure placement on or inside the envelope in a confidential way with an additional label, whereas the efficiency of this solution can be low. Hence, the label should act as an essential element of the letter, for example as prepaid envelope or the postal stamp dedicated for cross-border mail. Additionally, passive tags need to have relatively large antennas to achieve the appropriate radio coverage. For example, a typical passive RFID tag (complies with ISO 18000-6c) for coverage above 5 m needs antenna as big as 100×25 mm. Such a large component is difficult to hide. Another problem is that some senders may try to damage the tag. Embedding the tag on a postage stamp can solve this, but then it remains to solve the tag resistance to damage by timestamp hammer.

Neither of the mentioned solutions ensures end-to-end measurements, forcing an addition of a correction factor related to delay of collection and delivery. Delay of collection is relatively easy to estimate based on detailed records from domestic transit time. The estimation of delay of delivery is possible only with test letters with RFID tags.

Generally, all the proposed solutions generate significant costs for operators and do not cover all letters. The mentioned approaches do not ensure end-to-end measurements which means that all of them need adding a correction figure related to delay of collection and delivery. Hence, none of them seem to be ideal.

5. Conclusion

Postal items with correspondence are increasingly often being replaced by electronic communication, i.e. voice transmission or e-mail, thus decreasing the letters volume. Therefore, the cost of measurements of quality of service should also go down. However, the choice of the optimal solution will require further testing.

References

[1] Directive 97/67/EC of the European Parliament and of the Council of 15 December 1997 on common rules for the development of the internal market of Community postal services and the improvement of quality of service, OJ L 015 of 21 January 1998 with modifications (Directive 2002/39/EC and Directive 2008/6/EC).

[2] EN 13850:2002, "Postal services – Quality of service – Measurement of the transit time of end-to-end services for single piece priority mail and first class mail", 25 July 2002.

[3] R. Kobus, "Zastosowanie norm pocztowych", *Telekomunikacja i Techniki Informacyjne*, no. 1-2, pp. 67–81, 2012 (in Polish).

[4] International Post Corporation [Online]. Available: https://www.ipc.be/en/operational-services/quality%20excellence/unex

[5] R. Kobus, "Zastosowanie RFID do lokalizacji przesyłek pocztowych", *Zeszyty Naukowe Uniwersytetu Szczecińskiego nr 598, Ekonomiczne problemy usług nr 58*, pp. 657–664, 2010 (in Polish).

[6] EN 13850:2002+A1, "Postal service – Quality of service – Measurement of the transit time of end-to end-services for single piece priority mail and first class mail", March 2007.

[7] EN 13850:2012, "Postal Services – Quality of Services – Measurement of the transit time of end-to-end services for single piece priority mail and first class mail", 27 October 2012.

Ryszard Kobus received his B.Sc. and M.Sc. degrees from the Faculty of Electronics of the Warsaw University of Technology in 1975. Mr. Kobus has been working at the National Institute of Telecommunications since 1975. He is a member of the Expert Technical Committee CEN/TC 331 specializing in postal services, and the deputy chairman of the Postal Service Committee PKN/TC 259. He is a co-author of many patented telecommunications solutions. His research interests include: telecommunications, measurements and evaluation of quality of telecommunications services, quality surveys, evaluation the quality of postal services, standardization.
E-mail: R.Kobus@itl.waw.pl
National Institute of Telecommunications
1 Szachowa St
04-894 Warsaw, Poland

Frank Raudszus received his M.Sc. in Social Sciences and his Ph.D. in Economics at the University of Osnabrück. He continued with post graduate studies of "European Integration" at the University of Saarbrücken. Since 1991 he has been working for the Federal Network Agency. He has been chair of the DIN German Working Committee "Postal Services" and since 2016 he has chaired the European Technical Committee CEN/TC 331 Postal Services. Since 2011 he has also been working at the UPU Standards Board.
E-mail: Frank.Raudszus@BNetzA.de
Bundesnetzagentur
Canisiusstraße 21
Postfach 80 01, 55003 Mainz, Germany