

МСЭ-Т

СЕКТОР СТАНДАРТИЗАЦИИ
ЭЛЕКТРОСВЯЗИ МСЭ

E.840

(06/2018)

СЕРИЯ E: ОБЩАЯ ЭКСПЛУАТАЦИЯ СЕТИ,
ТЕЛЕФОННАЯ СЛУЖБА, ФУНКЦИОНИРОВАНИЕ
СЛУЖБ И ЧЕЛОВЕЧЕСКИЕ ФАКТОРЫ

Качество услуг электросвязи: концепции, модели, цели
и планирование надежности работы – Модели
для услуг электросвязи

**Статистическая основа для оценки
и ранжирования контрольных показателей
сквозных рабочих характеристик сети**

Рекомендация МСЭ-Т E.840

РЕКОМЕНДАЦИИ МСЭ-Т СЕРИИ E

ОБЩАЯ ЭКСПЛУАТАЦИЯ СЕТИ, ТЕЛЕФОННАЯ СЛУЖБА, ФУНКЦИОНИРОВАНИЕ СЛУЖБ
И ЧЕЛОВЕЧЕСКИЕ ФАКТОРЫ

МЕЖДУНАРОДНАЯ ЭКСПЛУАТАЦИЯ	
Определения	E.100–E.103
Общие положения, касающиеся администраций	E.104–E.119
Общие положения, касающиеся пользователей	E.120–E.139
Эксплуатация услуг международной телефонной связи	E.140–E.159
План нумерации для услуг международной телефонной связи	E.160–E.169
Международный план маршрутизации	E.170–E.179
Тональные сигналы в национальных системах сигнализации	E.180–E.189
План нумерации для услуг международной телефонной связи	E.190–E.199
Морская подвижная служба и сухопутная подвижная служба общего пользования	E.200–E.229
ЭКСПЛУАТАЦИОННЫЕ ПОЛОЖЕНИЯ, ОТНОСЯЩИЕСЯ К НАЧИСЛЕНИЮ ПЛАТЫ И РАСЧЕТАМ ЗА УСЛУГИ МЕЖДУНАРОДНОЙ ТЕЛЕФОННОЙ СВЯЗИ	
Начисление платы за услуги международной телефонной связи	E.230–E.249
Измерение и регистрация продолжительности разговоров в целях расчетов	E.260–E.269
ИСПОЛЬЗОВАНИЕ МЕЖДУНАРОДНОЙ ТЕЛЕФОННОЙ СЕТИ ДЛЯ ПРИЛОЖЕНИЙ, НЕ СВЯЗАННЫХ С ТЕЛЕФОННОЙ СВЯЗЬЮ	
Общие сведения	E.300–E.319
Фототелеграфная связь	E.320–E.329
ВОЗМОЖНОСТИ ЦСИС, ОТНОСЯЩИЕСЯ К ПОЛЬЗОВАТЕЛЯМ	E.330–E.349
МЕЖДУНАРОДНЫЙ ПЛАН МАРШРУТИЗАЦИИ	E.350–E.399
УПРАВЛЕНИЕ СЕТЬЮ	
Статистические данные по международным услугам	E.400–E.404
Управление международной сетью	E.405–E.419
Осуществление контроля качества услуг международной телефонной связи	E.420–E.489
ТЕХНИЧЕСКИЕ АСПЕКТЫ ТРАФИКА	
Измерение и регистрация трафика	E.490–E.505
Прогнозирование трафика	E.506–E.509
Определение количества каналов при ручном обслуживании	E.510–E.519
Определение количества каналов при автоматическом и полуавтоматическом обслуживании	E.520–E.539
Категория обслуживания	E.540–E.599
Определения	E.600–E.649
Технические аспекты трафика для IP-сетей	E.650–E.699
Технические аспекты трафика в ЦСИС	E.700–E.749
Технические аспекты трафика в сети подвижной связи	E.750–E.799
КАЧЕСТВО УСЛУГ ЭЛЕКТРОСВЯЗИ: КОНЦЕПЦИИ, МОДЕЛИ, ЦЕЛИ И ПЛАНИРОВАНИЕ НАДЕЖНОСТИ РАБОТЫ	
Термины и определения, связанные с качеством услуг электросвязи	E.800–E.809
Модели для услуг электросвязи	E.810–E.844
Показатели качества обслуживания и понятия, связанные с услугами электросвязи	E.845–E.859
Использование показателей качества обслуживания для планирования сетей электросвязи	E.860–E.879
Сбор эксплуатационных данных и оценка качества работы оборудования, сетей и служб	E.880–E.899
ДРУГИЕ	E.900–E.999
МЕЖДУНАРОДНАЯ ЭКСПЛУАТАЦИЯ	
План нумерации для услуг международной телефонной связи	E.1100–E.1199
УПРАВЛЕНИЕ СЕТЬЮ	
Управление международной сетью	E.4100–E.4199

Для получения более подробной информации просьба обращаться к перечню Рекомендаций МСЭ-Т.

Рекомендация МСЭ-Т E.840

Статистическая основа для оценки и ранжирования контрольных показателей сквозных рабочих характеристик сети

Резюме

Рекомендация МСЭ-Т E.840 – первая в серии Рекомендаций, посвященная эталонному (рейтинговому) тестированию сквозных рабочих характеристик сетей. В Рекомендации МСЭ-Т E.840 представлена основа для статистического анализа, который составляет базу для эталонного тестирования рабочих характеристик сетей и услуг. Основа включает в себя, в частности, сценарии эталонного тестирования, варианты использования, а также порядок и статистические методы ранжирования сквозных ключевых показателей работы (KPI) или ключевых показателей качества (KQI). Рекомендация МСЭ-Т E.840 относится к услугам подвижной связи и мероприятиям по эталонному тестированию, проводимым с использованием мобильных агентов (устройств) в условиях движения на автомобиле или ходьбы, а также фиксированных агентов (устройств), расположенных в фиксированных местах (например, в торговых центрах, офисных зданиях или на стадионах).

Хронологическая справка

Издание	Рекомендация	Утверждение	Исследовательская комиссия	Уникальный идентификатор*
1.0	МСЭ-Т E.840	13.06.2018 г.	12-я	11.1002/1000/13621

Ключевые слова

Сквозные рабочие характеристики, эталонное (рейтинговое) тестирование и ранжирование показателей работы сетей, статистическая основа.

* Для получения доступа к Рекомендации наберите в адресном поле вашего браузера URL: <http://handle.itu.int/>, после которого следует уникальный идентификатор Рекомендации. Например, <http://handle.itu.int/11.1002/1000/11830-en>.

ПРЕДИСЛОВИЕ

Международный союз электросвязи (МСЭ) является специализированным учреждением Организации Объединенных Наций в области электросвязи и информационно-коммуникационных технологий (ИКТ). Сектор стандартизации электросвязи МСЭ (МСЭ-Т) – постоянный орган МСЭ. МСЭ-Т отвечает за изучение технических, эксплуатационных и тарифных вопросов и за выпуск Рекомендаций по ним в целях стандартизации электросвязи на всемирной основе.

На Всемирной ассамблее по стандартизации электросвязи (ВАСЭ), которая проводится каждые четыре года, определяются темы для изучения исследовательскими комиссиями МСЭ-Т, которые, в свою очередь, вырабатывают Рекомендации по этим темам.

Утверждение Рекомендаций МСЭ-Т осуществляется в соответствии с процедурой, изложенной в Резолюции 1 ВАСЭ.

В некоторых областях информационных технологий, которые входят в компетенцию МСЭ-Т, необходимые стандарты разрабатываются на основе сотрудничества с ИСО и МЭК.

ПРИМЕЧАНИЕ

В настоящей Рекомендации термин "администрация" используется для краткости и обозначает как администрацию электросвязи, так и признанную эксплуатационную организацию.

Соблюдение положений данной Рекомендации осуществляется на добровольной основе. Однако данная Рекомендация может содержать некоторые обязательные положения (например, для обеспечения функциональной совместимости или возможности применения), и в таком случае соблюдение Рекомендации достигается при выполнении всех указанных положений. Для выражения требований используются слова "следует", "должен" (shall) или некоторые другие обязывающие выражения, такие как "обязан" (must), а также их отрицательные формы. Употребление таких слов не означает, что от какой-либо стороны требуется соблюдение положений данной Рекомендации.

ПРАВА ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

МСЭ обращает внимание на вероятность того, что практическое применение или выполнение настоящей Рекомендации может включать использование заявленного права интеллектуальной собственности. МСЭ не занимает какую бы то ни было позицию относительно подтверждения, действительности или применимости заявленных прав интеллектуальной собственности независимо от того, доказываются ли такие права членами МСЭ или другими сторонами, не относящимися к процессу разработки Рекомендации.

На момент утверждения настоящей Рекомендации МСЭ не получил извещения об интеллектуальной собственности, защищенной патентами, которые могут потребоваться для выполнения настоящей Рекомендации. Однако те, кто будет применять Рекомендацию, должны иметь в виду, что вышесказанное может не отражать самую последнюю информацию, и поэтому им настоятельно рекомендуется обращаться к патентной базе данных БСЭ по адресу <http://www.itu.int/ITU-T/ipr/>.

© ITU 2020

Все права сохранены. Ни одна из частей данной публикации не может быть воспроизведена с помощью каких бы то ни было средств без предварительного письменного разрешения МСЭ.

СОДЕРЖАНИЕ

	Стр.
1 Сфера применения	1
2 Справочные документы	1
3 Определения	1
4 Сокращения и акронимы	1
5 Условные обозначения	1
6 Сценарии эталонного (рейтингового) тестирования	2
7 Условия эталонного (рейтингового) тестирования.....	3
8 Услуги, подлежащие эталонному (рейтинговому) тестированию	4
9 Статистическая основа	5
9.1 Очистка данных	5
9.2 Статистическое распределение измерительных данных	5
9.3 Статистические метрики рабочих характеристик, стандартная ошибка и статистическая значимость результатов эталонного тестирования.....	5
9.4 Оценка и ранжирование сквозных KPI или KQI	7
Приложение А – Уровни статистической значимости, применяемые при сравнительном анализе рабочих характеристик сетей подвижной связи	9
Приложение В – Статистическая оценка и ранжирование рабочих характеристик сети	10
Дополнение I – Возможная методика статистической оценки и ранжирования сети	12
Библиография	14

Рекомендация МСЭ-Т E.840

Статистическая основа для оценки и ранжирования контрольных показателей сквозных рабочих характеристик сети

1 Сфера применения

В настоящей Рекомендации представлена статистическая основа, которую предлагается использовать операторам и регуляторным органам при качественной и количественной оценке различий в сквозных ключевых показателях работы (KPI) или ключевых показателях качества (KQI), затрагивающих опыт пользователя, а также сценарии и условия эталонного (рейтингового) тестирования, в которых эта основа может быть применена.

Потребность в этой Рекомендации возникла в связи с тем, что в стремлении удовлетворить растущие запросы существующих пользователей и одновременно расширить клиентскую базу при оптимальных затратах операторы повысили качество работы своих сетей до такого уровня, при котором различия между ними становятся все менее существенными.

2 Справочные документы

Указанные ниже Рекомендации МСЭ-Т и другие справочные документы содержат положения, которые путем ссылок на них в данном тексте составляют положения настоящей Рекомендации. На момент публикации указанные издания были действующими. Все Рекомендации и другие источники могут подвергаться пересмотру; поэтому всем пользователям данной Рекомендации предлагается изучить возможность применения последнего издания Рекомендаций и других справочных документов, перечисленных ниже. Список действующих в настоящее время Рекомендаций МСЭ-Т регулярно публикуется. Ссылка на документ в данной Рекомендации не придает ему как отдельному документу статус Рекомендации.

[ITU-T E.800] Рекомендация МСЭ-Т E.800 (2008 г.), *Определение терминов, относящихся к качеству обслуживания.*

[ITU-T E.804] Recommendation ITU-T E.804 (2014), *QoS aspects for popular services in mobile networks.*

3 Определения

Отсутствуют.

4 Сокращения и акронимы

В настоящей Рекомендации используются следующие сокращения и акронимы.

KPI	Key Performance Indicator		Ключевой показатель работы
KQI	Key Quality Indicator		Ключевой показатель качества
QoE	Quality of Experience		Оценка пользователем качества услуги
MOS	Mean Opinion Score		Средняя экспертная оценка
RF	Radio Frequency	РЧ	Радиочастота
TCP	Transmission Control Protocol		Протокол управления передачей

5 Условные обозначения

5.2.1 StatScore: Статистическая оценка относительного общего качества работы различных сетей или операторов по сравнению с сетью, демонстрирующей наилучшие показатели работы. Рассчитывается для каждой услуги.

5.2.2 GlobalNetScore: Глобальная оценка сети, то есть оценка относительного общего качества работы различных сетей или операторов по сравнению с сетью, демонстрирующей наилучшие показатели работы. Рассчитывается по всем услугам.

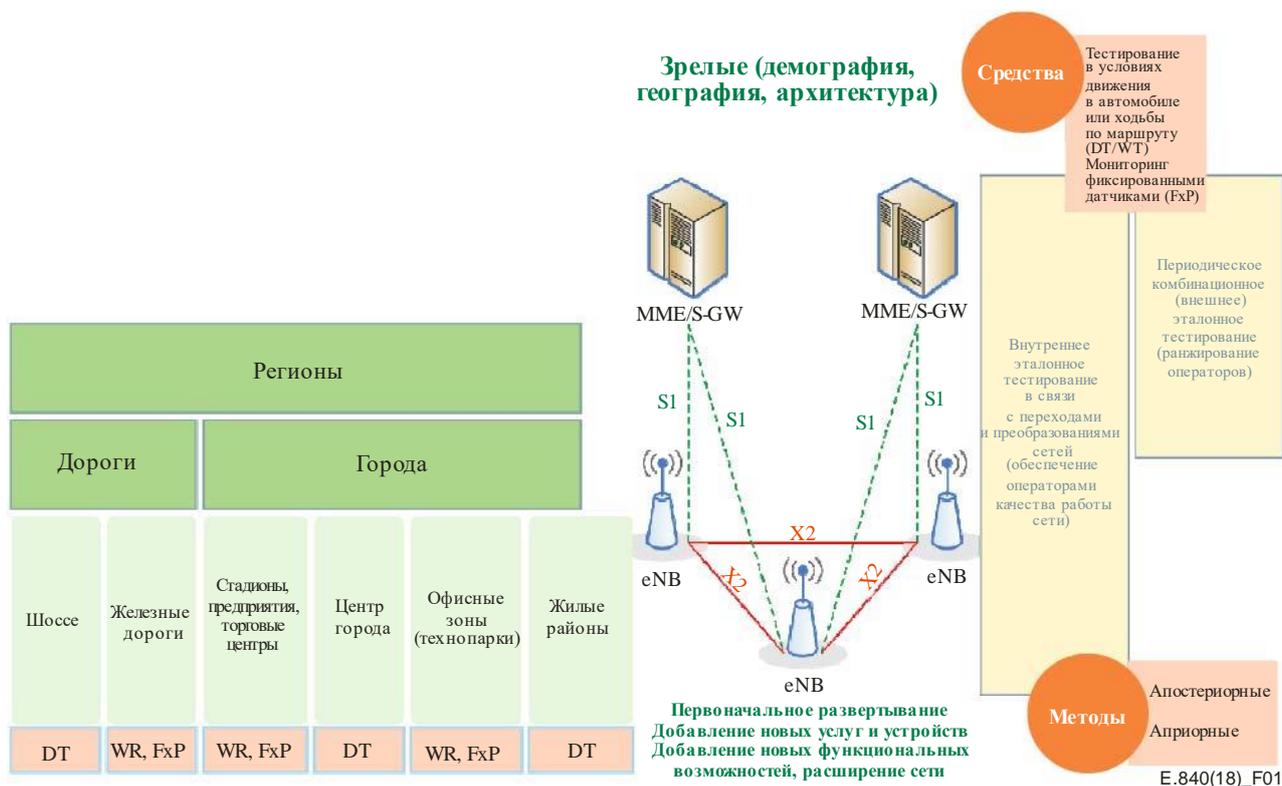
5.2.3 StatDiff: Статистически значимое различие между двумя сравниваемыми ключевыми показателями работы (KPI) или ключевыми показателями качества (KQI).

5.2.4 THrelv: Минимальное различие между значениями двух конкретных ключевых показателей работы (KPI) или ключевых показателей качества (KQI), которое субъективно значимо с точки зрения пользователя услуги и которое по приоритетности уступает статистической значимости.

6 Сценарии эталонного (рейтингового) тестирования

Эталонное тестирование сетей обычно используется в двух вариантах – внутреннем и конкурентном. При внутреннем эталонном тестировании оценивается непрерывное экономически эффективное обеспечение качества работы сети, а также усовершенствования, требующие оценки при первоначальном развертывании сети, в ходе ее разработки и при выводе на рынок новых услуг и устройств. Внутреннее эталонное тестирование проводится также в устоявшихся и зрелых сетях. Кроме того, в ходе мероприятий по оценке необходимо рассматривать регионы с шоссейными дорогами и города, а также конкретные зоны, представляющие интерес (например, рабочие места, торговые центры, стадионы, жильё). Конкурентное эталонное тестирование, осуществляемое самими операторами (или компаниями – поставщиками услуг от имени операторов), а также регуляторными органами для оценки конкурентов и саморанжирования, обычно проводится по некоторому множеству регионов, транспортных маршрутов (шоссейных, железных дорог), городов и даже стран (в случае многонациональных групп операторов и зрелых сетей).

Сводная информация о вариантах использования, а также рекомендуемых средствах и методах представлена на рисунке 1. Эталонное тестирование конкретных зон (например, торговых центров, стадионов и рабочих мест, обычно в помещениях) часто проводят в условиях ходьбы. Помимо традиционных мероприятий по тестированию в условиях движения в автомобиле или ходьбы по определенному маршруту, при проведении внутреннего и до некоторой степени конкурентного эталонного тестирования, а также тестирования в помещениях полезно применять фиксированные средства на базе датчиков. Их преимуществами являются быстрое и дистанционное масштабирование, а также аппаратная независимость. Таким образом эти средства хорошо подходят для тестирования в помещениях и проведения мероприятий по выводу на рынок новых услуг в представляющих интерес зонах, а до некоторой степени и в городах. Кроме того, можно видеть, что могут применяться апостериорные или априорные методы анализа. Апостериорные методы (применяемые главным образом в эталонном тестировании) предусматривают, что после сбора данных производится оценка и ранжирование сквозных KPI или KQI на основании статистической значимости; точность измерения по умолчанию заложена в уровень статистической значимости. Априорные методы предполагают предварительный расчет числа тестовых датчиков, необходимых для достижения заданного уровня статистической значимости и точности измерений. Такие методы обычно применяются, когда тестовые датчики дороги или время, выделенное на тестирование, ограничено.



E.840(18)_F01

Рисунок 1 – Варианты использования, рекомендуемые средства и методы эталонного тестирования

7 Условия эталонного (рейтингового) тестирования

Вне зависимости от варианта использования основа эталонного тестирования должна опираться на соблюдение ряда необходимых требований, обеспечивающих единообразие, корректность, надежность и повторяемость. Эти требования для каждого из этапов эталонного тестирования (состав оборудования, конфигурация теста, сбор данных, обработка и анализ данных) приведены в таблице 1. Следует отметить, что в таблице 1 приведены минимальные требования, необходимые для обеспечения полностью контролируемых условий тестирования, а также корректности статистического анализа. Подробное описание соответствующих измерений дается в других Рекомендациях МСЭ-Т (например, [ITU-T E.804]).

Таблица 1 – Минимально необходимые требования

Эталонное тестирование	Необходимые требования к эталонному тестированию
Состав оборудования	Состав оборудования должен быть единообразным для различных сетей, платформ и устройств; для конкурентного эталонного тестирования должны использоваться одни и те же модели устройств (в целях сравнения "подобного с подобным").
	Оборудование должно работать в надлежащих условиях, предписанных поставщиком (например, необходимо избегать перегрева, так как он может негативно сказаться на работе устройства).

Таблица 1 – Минимально необходимые требования

Эталонное тестирование	Необходимые требования к эталонному тестированию
Схема и конфигурация тестовой установки	Схема теста должна отражать реальный опыт пользователя; руководящие указания на этот счет можно найти в [b-ETSI TR 102 581].
	Параметры тестовой установки должны быть такими, чтобы избежать артефактов, которые могут искусственно повлиять на качество работы сети; например, следует проверить расположение серверов данных и параметры протокола TCP, чтобы обеспечить надлежащую пропускную способность для всех сравниваемых операторов. Этот тест следует выполнить перед тем, как приступить к сбору данных.
	Разработка различных сценариев тестирования, отражающих поведение пользователей (например, использование протокола TCP, разные размеры файлов, разная длительность видео- и голосовых вызовов), но с наименьшей возможной инвазивностью, чтобы искусственно не перегружать сеть.
Сбор данных	Необходимо собирать данные, отражающие опыт пользователя [например, среднюю экспертную оценку (MOS) каждой услуги, а также основные сквозные сетевые KPI или KQI, влияющие на нее]. В основу выполняемых измерений должны быть положены адекватные метрики или методы в соответствии с руководящими указаниями поставщиков или положениями соответствующих Рекомендаций.
	Собирать данные необходимо в различных географических или демографических условиях и в разные характерные периоды времени (часы пиковой и обычной нагрузки, выходные и будни, отпускной и рабочий сезон).
Обработка и анализ данных	Сравнивать необходимо данные, собранные с использованием одних и тех же устройств, в одной и той же зоне, в течение одного и того же характерного периода времени (в целях сравнения "подобного с подобным").
	Необходимо применять критерии статистической значимости для корректного сравнения.
	Анализ должен выполняться по каждому KPI или KQI.

8 Услуги, подлежащие эталонному (рейтинговому) тестированию

Перечень услуг подвижной связи, обычно подлежащих эталонному тестированию, их KPI или KQI и условий приведения их в действие выходит за рамки настоящей Рекомендации. Подробнее см. в [ITU-T E.800] и [ITU-T E.804].

Если задачей эталонного тестирования услуг подвижной связи является проведение детального сравнительного анализа по каждой услуге, что обычно осуществляется при внутреннем эталонном тестировании вариантов использования (например, при добавлении нового устройства или новой технологии), то для анализа рекомендуется использовать комплексный набор сквозных KPI и KQI (здесь под KQI понимаются величины, получаемые с использованием моделей оценки качества, например [b-ITU-T P.863] для голосовой связи или [b-ITU-T P.1203] для потоковой передачи видео). Вдобавок на основании этого набора рекомендуется проводить анализ основных коренных причин возможного неудовлетворительного уровня рабочих характеристик.

Если же целью эталонного тестирования услуг подвижной связи является ранжирование рабочих характеристик по сквозным KPI или KQI, как обычно осуществляется при конкурентном эталонном тестировании вариантов использования и в некоторых сценариях внутреннего эталонного тестирования (например, сравнение с рынком, периодическая оценка рыночной конъюнктуры), то для целей оценки и ранжирования можно рассматривать более узкий набор KPI или KQI, влияющих на оценку качества пользователем (QoE; см. [b-ITU-T P.10/G.100]), по каждой услуге и по всем тестируемым услугам.

В настоящей Рекомендации предполагается последний вариант конкурентного сравнительного анализа. Подробное описание наборов таких KPI или KQI дается в других Рекомендациях МСЭ-Т.

9 Статистическая основа

Рекомендуемая здесь статистическая основа предназначена для оценки и ранжирования сквозных рабочих характеристик сети с точки зрения пользователя и может применяться как для конкурентного, так и для внутреннего эталонного тестирования. Она определяет порядок проверки корректности данных, метрики статистической оценки и критерии статистической значимости, а также содержит общие руководящие принципы ранжирования и оценки.

9.1 Очистка данных

Для обеспечения корректности и точности результатов эталонного тестирования необходимо проверять корректность входных данных, используемых для анализа. Проверка корректности в основном сводится к очистке данных, то есть исключению артефактов измерения и неполных данных. Отсутствующие данные рекомендуется заменять собранными в последнее время. Если отсутствует 5% данных и более, рекомендуется собрать новые данные. К числу артефактов измерения относятся, среди прочего, "молчаливые" вызовы в случае услуг голосовой или видеосвязи, периоды недоступности серверов данных или видео (например, по причине выключения сервера) в случае услуг потоковой передачи видео и систематически возникающие против ожидания сверхнизкие значения MOS в случае голосовой связи, видеосвязи или потоковой передачи видео. Здесь любое снижение качества, вызванное не сетью как таковой, а тестовым оборудованием или устройством, следует расценивать как артефакт измерения.

Данные эталонного тестирования необходимо очистить от таких артефактов, чтобы обеспечить корректность предположений, сделанных при применении статистических моделей.

9.2 Статистическое распределение измеренных данных

Вообще говоря, согласно центральной предельной теореме [b-Shaum], статистическое распределение значений любых измеренных KPI или KQI можно аппроксимировать гауссовым (нормальным) распределением. Как правило, чем объемнее выборка, тем выше точность такой аппроксимации. Рекомендуется проверять статистические распределения всех значений KPI или KQI. Делать это можно двумя способами: либо строить графики распределений измеренных значений, анализируемых KPI или KQI, и проверять их нормальность путем визуального наблюдения, либо применять для проверки нормальности критерии согласия Колмогорова–Смирнова, Андерсона–Дарлингга или Шапиро–Уилка [b-Mehta]. В тех редких или крайних случаях, когда приходится иметь дело с негауссовыми распределениями, можно применять непараметрические тесты. Тесты, демонстрирующие полученные экспериментально распределения различных KPI и KQI, тесты для проверки нормальности распределения, а также особые случаи негауссовых распределений рассматриваются в других Рекомендациях МСЭ-Т из серии, посвященной эталонному тестированию.

9.3 Статистические метрики рабочих характеристик, стандартная ошибка и статистическая значимость результатов эталонного тестирования

9.3.1 Статистические метрики рабочих характеристик

В основу сравнительного анализа должны быть положены статистические метрики рабочих характеристик, отражающие усредненные характеристики сети (представленные средними значениями m) или их постоянство (представленное вероятностью P_{th} превышения заданного порогового значения). В настоящей Рекомендации для примера рассматривается статистическая метрика среднего значения рабочих характеристик. Аналогичные методы могут применяться для анализа постоянства.

9.3.2 Стандартная ошибка

Стандартную ошибку среднего или P_{th} при доверительной вероятности 95% рассчитывают в предположении нормального распределения измеряемых KPI или KQI (см. пункт 9.2).

Соответственно в зависимости от типа KPI или KQI, непрерывно оцениваемых показателей (например, MOS или радиочастотных параметров) либо оцениваемых дискретно (например, отношение r числа успешных и неудачных исходов), стандартная ошибка при доверительной вероятности 95% дается формулами:

$$\text{StdError}(m) = z_{95\%} * \text{std}/\sqrt{N} = 1,96 * \text{std}/\sqrt{N};$$

$$\text{StdError}(r) = z_{95\%} * \sqrt{r*(1-r)/N} = 1,96 * \sqrt{r*(1-r)/N}.$$

Если в выборке менее 30 элементов, квантиль нормального распределения $z_{95\%}$ следует заменить табличным значением коэффициента Стьюдента $t_{95\%}(N-1)$, где N – число элементов в выборке.

ПРИМЕЧАНИЕ. – Стандартная ошибка является показателем точности измерения. Поэтому если требуется конкретный уровень точности, а оценка стандартного отклонения известна, то можно определить минимально необходимое число элементов в выборке для достижения данного уровня точности при выбранной доверительной вероятности, используя формулы во втором абзаце. Этот подход можно использовать как *априорный* метод (см. рисунок 1), и он используется в [b-ITU-T E.802] для расчета минимального объема выборки.

9.3.3 Статистическая значимость

Доверительный интервал задает границы стандартной ошибки при заданном уровне доверительной вероятности (обычно 95%). Однако два сравниваемых KPI или KQI с близкими значениями и перекрывающимися доверительными интервалами не обязательно представляют собой одно и то же со статистической точки зрения. Точное сравнение KPI или KQI должно основываться на статистической значимости. Тем самым вероятность ошибочно отвергнуть верную гипотезу об идентичности значений двух KPI или KQI удерживается на уровне 5%.

Критерий статистической значимости позволяет заключить, что один оператор работает лучше другого (при конкурентном эталонном тестировании вариантов использования), а также определить, является ли новая технология или функциональная возможность причиной существенного улучшения (при внутреннем эталонном тестировании).

Помимо критериев статистической значимости необходимо использовать пороги значимости различий, установленные для конкретных KPI или KQI, в тех случаях, когда различия несущественны или находятся в пределах точности измерения соответствующего KPI или KQI. Пороги значимости различий для конкретных KPI или KQI (THrelv) определены в других Рекомендациях МСЭ-Т из серии, посвященной эталонному тестированию.

Пример того, как необходимо сравнивать KPI или KQI, приведен в таблице 2. Для обеспечения общего характера настоящей Рекомендации в качестве примеров для сравнения взяты условные метрики KQI1 и KQI2. Таким образом в роли KQI1 и KQI2 могут выступать любые KQI (или KPI), выбранные для конкретной услуги. Сначала вычисляются выборочное среднее и стандартное отклонение KQI1 и KQI2. Исходя из объема имеющейся выборки и применяя критерии A-1 и A-2, можно заключить, что качество, обеспечиваемое сетями 1 и 2 по метрикам KQI1 и KQI2, статистически одинаково. Кроме того, можно видеть, что различия по каждому из KQI (KQI1 и KQI2) остаются на уровне ниже THrelv. Статистически значимая разница в рабочих характеристиках сетей 1 и 2 наблюдается по третьей метрике – KQI3. Тем не менее размер этой разницы (0,02) остается ниже THrelv (0,025), поэтому вывод о различии рабочих характеристик сетей 1 и 2 сделать нельзя.

Таблица 2 – Пример применения критерия статистической значимости

Услуга	KQI	Сеть 1			Сеть 2			Статистика при доверит. вероятности 95% (Z > 1,96)	StatDiff	THrelv
		Среднее	std	N	Среднее	std	N			
Услуга 1	KQI1	3,27	0,3	287	3,35	0,6	212	1,78	Нет	0,09
	KQI2	0,02	0,14	12	0,015	0,12	10	0,09	Нет	0,006
	KQI3	0,93	0,26	69	0,91	0,29	71	2,04	Да	0,025

Такого рода анализ результатов подробного эталонного тестирования можно расширенно применять для различных услуг, а также для более объемного набора KPI или KQI на одну услугу, как отмечалось в разделе 8. Кроме того, по результатам проверки статистической значимости (столбец "статистика при доверительной вероятности 95%", таблица 2) можно ранжировать сети по отдельным KPI или KQI, как описано в пункте 9.4.

Необходимо отметить следующее: чтобы утверждать, что одну сеть или конфигурацию услуги можно считать лучшей в сравнении с другой сетью или конфигурацией, необходимо, помимо проверки статистической значимости, определить порог значимых различий для конкретного KPI или KQI и предоставить информацию о точности измерения, как определено в других Рекомендациях МСЭ-Т из серии, посвященной эталонному тестированию.

9.3.4 Протоколирование результатов

Результаты статистического анализа эталонного тестирования, а также подробное описание сценариев и условий тестирования необходимо запротоколировать, в противном случае результаты могут быть интерпретированы неверно и, соответственно, будут некорректны.

9.4 Оценка и ранжирование сквозных KPI или KQI

Сквозные KPI и KQI сетей могут быть оценены и ранжированы по каждому из аспектов, рассматриваемых в рамках мероприятий по эталонному тестированию. Кроме того, как уже отмечалось в пункте 9.3.3, чтобы обнаружить все более тонкие различия в рабочих характеристиках сетей действующих операторов, следует исходить из статистической значимости. В тех случаях, когда различия несут существенны или находятся в пределах точности измерения, необходимо пользоваться порогами значимости различий для соответствующих KPI или KQI, как описано в других Рекомендациях МСЭ-Т из серии, посвященной эталонному тестированию.

Методика статистической балльной оценки и ранжирования по набору KPI или KQI приведена в приложении В, а пример приведен в таблице 3.

Таблица 3 – Пример оценки и ранжирования сквозных KPI или KQI с учетом статистической значимости

	Сеть 1				Сеть 2				
	KPI/KQI	std	N	StatDiff	KPI/KQI	std	N	StatDiff	THrelv
KPI1/KQI1	0,95	0,22	87	0,05	0,97	0,17	69	0,00	0,018
KPI2/KQI2	0,93	0,26	87	0,00	0,91	0,29	69	0,23	0,019
KPI3/KQI3	3,89	0,50	2 600	0,00	3,56	0,70	2 070	17,15	0,31
KPI4/KQI4	105,00	5,00	435	42,67	70,00	15,00	350	0,00	34
KPI5/KQI5	1 200,00	300,00	87	0,00	1 800,00	275,00	69	12,31	596

В таблице 3 приведены агрегированные значения, которые были получены в одной и той же зоне на протяжении одного и того же характерного периода времени, чтобы сравнивать "подобное с подобным" (см. руководящие указания в таблице 1). Помимо значений рабочих характеристик, приведены расчетные значения стандартного отклонения и объема выборки. Для каждого KPI или KQI приведено также расчетное значение параметра StatDiff (см. Приложение В) как статистически значимое расхождение с наилучшим достигнутым значением данного KPI или KQI (выделено желтым в таблице 3) при доверительной вероятности 95% (см. Приложение В). Чем ниже StatDiff, тем ближе значение KPI или KQI к наилучшему из достигнутых; StatDiff, равное 0, представляет наилучшее значение KPI или KQI. Кроме того, можно отметить, что во всех случаях различия в KQI двух сетей превышают пороговый уровень THrelv, то есть приоритет всегда у статистической значимости.

На основании этих данных в таблице 4 проведено ранжирование значений KPI или KQI.

Таблица 4 – Пример статистического ранжирования значений KPI или KQI

KPI/KQI	Сеть 1	Сеть 2
KPI1/KQI1	Ранг 2	Ранг 1
KPI2/KQI2	Ранг 1	Ранг 2
KPI3/KQI3	Ранг 1	Ранг 2
KPI4/KQI4	Ранг 2	Ранг 1
KPI5/KQI5	Ранг 1	Ранг 2

Это ранжирование можно распространить и на более широкий набор KPI или KQI, который может применяться для подробного эталонного тестирования или внутреннего тестирования вариантов использования.

В некоторых случаях может потребоваться статистическая оценка сети по каждой услуге и по всем услугам. В Приложении А описывается возможная методика такой оценки. Следует однако отметить, что подобная методика будет корректной только в том случае, если она полностью описана и базируется на технически обоснованных предположениях и условиях.

Приложение А

Уровни статистической значимости, применяемые при сравнительном анализе рабочих характеристик сетей подвижной связи

(Данное Приложение является неотъемлемой частью настоящей Рекомендации.)

Сравнительный анализ предполагает сравнение KPI или KQI, описывающих рабочие характеристики сетей различных операторов. Корректное сравнение производится по итогам проверки статистической значимости (проверки гипотезы), которые зависят от типа сравниваемых KPI или KQI – непрерывных (например, MOS или радиочастотные параметры) либо дискретных (например, относительное число успешных или неудачных исходов).

В первом случае существенное различие определяется неравенством (A-1) [ITU-T P.1401]:

$$Z = \text{StatDiff}/\sqrt{\text{std1}^2/N1 + \text{std2}^2/N2} > Z_{th}, \quad (\text{A-1})$$

где StatDiff обозначает разность значений сравниваемых метрик, std1 и std2 – их стандартные отклонения, а N1 и N2 – полные объемы выборок, использовавшихся для сравнения по каждой метрике. Иными словами, если Z больше Zth (в предположении нормального распределения при объеме выборки более 30 элементов и доверительной вероятности CL%), то StatDiff представляет собой статистически значимую разность при данной доверительной вероятности.

Во втором случае тип соотношения KPI или KQI описывается числом p успешных или неудачных исходов из общего числа элементов в выборке. Статистическая значимость тогда определяется неравенством (A-2) [b-ITU-T P.1401]:

$$Z = \text{StatDiff}/\sqrt{p1*(1 - p1)/N1 + p2*(1 - p2)/N2} > Z_{th}, \quad (\text{A-2})$$

где p1 и p2 – число успешных или неудачных исходов по каждой из сравниваемых метрик.

В таблице А.1 показано соответствие между значениями порога значимости Zth и уровнями доверительной вероятности.

Таблице А.1 – Соответствие между значениями порога значимости и уровнями доверительной вероятности

CL %	90	95	96	97	98	99
Zth	1,64	1,96	2,05	2,17	2,33	2,58

Если в выборке менее 30 элементов, то следует использовать t-распределение Стьюдента t_95% (n), где n = N – 1 – число степеней свободы, а N – число элементов в выборке.

Следует отметить, что в тех случаях, когда различия незначительны или находятся в пределах точности измерения соответствующего KPI или KQI, помимо критериев статистической значимости, необходимо использовать соответствующие пороги значимости различий, установленные для конкретных KPI или KQI. Пороги значимости различий для конкретных KPI или KQI (THrelv) определены в других Рекомендациях МСЭ-Т из серии, посвященной эталонному тестированию.

Приложение В

Статистическая оценка и ранжирование рабочих характеристик сети

(Данное Приложение является неотъемлемой частью настоящей Рекомендации.)

В настоящем Приложении описан алгоритм оценки и ранжирования сквозных рабочих характеристик сетей, использованный для расчетов в таблице 3.

- Рассчитать сквозные KPI (или KQI) анализируемой услуги для каждой сети или оператора:
 - KPI_1...KPI_i...KPI_N, $i = 1, n$ (могут быть средними, медианными значениями или отношениями).
- Построить матрицу сравнения для $j = 1, M$ сетей (операторов) на каждую услугу, описываемую N метриками (KPI или KQI), см. таблицу В.1.

Таблица В.1 – Матрица сравнения

	Netwk_1.....	Netwk_j.....	Netwk_M
KPI_1...	KPI_1,1	KPI_1,j	KPI_1,M
KPI_i..	KPI_i,1	KPI_i,j	KPI_i,M
KPI_N	KPI_N,1	KPI_N,j	KPI_N,M

- Рассчитать расстояние статистической значимости для каждого элемента матрицы KPI_i,j.
Для каждого $i = 1, N$
 - выбрать наилучшее значение метрики KPI_i,best, где best – сеть j с наилучшими рабочими характеристиками;
 - рассчитать статистически значимую разность StatDiff_i,j для каждого KPI_i,j с наилучшим достигнутым значением KPI_i,best по формулам (В-1) и (В-2) в зависимости от типа метрики:

$$\text{StatDiff}_{i,j} = \max \{0, (KPI_{i,best} - KPI_{i,j}) / \sqrt{(\text{std}1^2/N1 + \text{std}2^2/N2)} - Z_{th}\}; \quad (\text{В-1})$$

$$\text{StatDiff}_{i,j} = \max \{0, (KPI_{i,best} - KPI_{i,j}) / \sqrt{p1*(1 - p1)/N1 + p2*(1 - p2)/N2)} - Z_{th}\}. \quad (\text{В-2})$$

ПРИМЕЧАНИЕ. – $Z_{th} = F(0,05, N1, N2)$ – статистический результат при доверительной вероятности 95% с $N1$ и $N2$ степенями свободы.

Конец

- Определить сеть или оператора с рангом 1 (наилучшие показатели работы) по тестируемой услуге в данной зоне.

Ранг 1 присваивается сети с наименьшим статистически значимым расстоянием по качеству тестируемой услуги среди множества сетей или операторов.

Для каждого $j = 1, M$

$$\text{Rank } 1 = \text{Rank}(j), \text{ если } \text{StatDiffQuality}_{\min} = \min(j=1, M) \{ \text{SUM}(i=1, N) \{ \text{StatDiff}_{i,j} * w_i \} \},$$

где StatDiff_i,j соответствует определению в формуле (В-1) или (В-2), а w_i представляет заранее определенные веса каждой услуги, если требуется учет весов. В противном случае можно использовать равные единичные веса.

Конец

Определить ранги всех сетей или операторов, рассматриваемых в рамках мероприятия по эталонному тестированию.

Для каждого $j = 1, M$

$$\text{Если } \text{Dist}(j) = \max(0, \text{StatDiffQuality} / \text{StatDiffQuality}_{\min} - Z_{th}) = 0,$$

$$\text{Rank}(j) = \text{Rank } 1$$

(определяет все сети с такими же статистическими показателями, как и у сети с рангом 1).

Иначе

Упорядочить $\text{Dist}(j)$ по возрастанию.

$\text{Rank}(j)$ = позиция в векторе $\text{Dist}(j)$.

Конец

Следует отметить, что в тех случаях, когда различия несущественны или находятся в пределах точности измерения соответствующего КРІ или КQІ, для ранжирования необходимо учитывать не только статистическую значимость, но и пороги значимости различий, установленные для конкретных КРІ или КQІ. Пороги значимости различий для конкретных КРІ или КQІ (Threlv) определены в других Рекомендациях МСЭ-Т из серии, посвященной эталонному тестированию.

Дополнение I

Возможная методика статистической оценки и ранжирования сети

(Это Дополнение не является неотъемлемой частью настоящей Рекомендации.)

Иногда бывает желательно получить статистическую оценку для всей сети в целом. В качестве базового критерия для этой цели часто используют показатели работы сети по каждому типу услуги. Такую оценку следует получить как для каждой зоны, так и по всем регионам вместе (рисунок 1).

Статистическую оценку по отдельной услуге можно определить на основании всех рассматриваемых сквозных метрик KPI_i или KQI_i (i = 1,N), влияющих на сквозное качество данной услуги, как взвешенную сумму StatDiff_i (см. Приложение B) по каждому KQI в сравнении с наилучшим достигнутым значением, как описано в пункте 9.4. Если установлен порог значимости различий, в значение StatDiff_i вносится соответствующая поправка. Результирующее значение StatScore описывает сквозные показатели работы рассматриваемых сетей в сравнении с сетью, демонстрирующей наилучшие показатели:

$$\text{StatScore} = \sum(w_i * \text{StatDiff}_i).$$

Здесь w_i – это вес каждой метрики KPI или KQI, вносящей вклад в качество услуги. Чем ниже StatScore, тем выше (ближе к показателям наилучшей сети) показатели работы и ранг.

Таблица I.1 представляет собой новую версию таблицы 3, в которую добавлены веса; они приведены только для примера, поскольку определение весов выходит за рамки настоящей Рекомендации. Веса могут быть унифицированными, если будет решено, что все KPI или KQI одинаково важны для общей статистической оценки показателей работы сети. Следует однако отметить, что даже если использовать унифицированные веса, при увеличении или уменьшении числа KPI или KQI статистическая оценка сети может измениться и дать другие статистические результаты.

Поэтому результаты статистической оценки и ранжирования сети согласно настоящей Рекомендации будут корректными только при условии подробного описания и мотивировки выбора и присвоения выбранных весов метрикам KPI или KQI. При отсутствии подобной прозрачности статистическая оценка и ранжирование сети не являются корректными.

Например, в таблице I.1 сети 1 присваивается высший ранг 1 с минимальной оценкой 2,15, исходя из присвоенных весов. Кроме того, можно отметить, что во всех случаях различия в KPI или KQI двух сетей превышают пороговый уровень THrelv, что означает корректность результатов анализа статистической значимости.

Таблица I.1 – Пример статистической оценки и ранжирования

	Сеть 1				Сеть 2				
	KPI	std	N	StatDiff	KPI	std	N	StatDiff	THrelv
KPI1/KQI1	0,95	0,22	87	0,05	0,97	0,17	69	0,00	0,018
KPI2/KQI2	0,93	0,26	87	0,00	0,91	0,29	69	0,23	0,019
KPI3/KQI3	3,89	0,50	2 600	0,00	3,56	0,70	2 070	17,15	0,31
KPI4/KQI4	105,00	5,00	435	42,67	70,00	15,00	350	0,00	34
KPI5/KQI5	1 200,00	300,00	87	0,00	1 800,00	275,00	69	12,31	596
StatScore				2,15				5,83	
Ранг				1				2	

В рамках комплексного мероприятия по эталонному тестированию может быть поставлена цель – рассчитать глобальную статистическую оценку для каждой из тестируемых сетей по всем поддерживаемым устройствам (j = 1,M; M – число поддерживаемых устройств). Следует отметить, что для обеспечения соответствия требованиям настоящей Рекомендации любая такая глобальная оценка должна иметь основательную мотивировку, а присваивание весов должно осуществляться явным и прозрачным образом.

Такого рода оценку можно рассчитать, сложив все статистические оценки по каждой услуге, рассчитанные согласно Приложению В (пример см. в таблице I.1). В таком случае у сети с наилучшими показателями работы значение глобальной статистической оценки равно нулю. Чем ниже оценка, тем лучше показатели работы соответствующей сети по результатам суммирования с весами KPI и KQI, установленные для каждой услуги. Оценка, получившаяся в результате такой операции, часто используется для глобального ранжирования сетей в сравнении с наилучшей:

$$\text{GlobalNetScore} = \sum(W_{\text{serv}_i} * \text{StatScore}_j).$$

Определение весов W_{serv_j} для GlobalNetScore выходит за рамки настоящей Рекомендации, но есть некоторые руководящие указания.

В зависимости от политики оператора или приоритетов бизнеса GlobalNetScore может рассчитываться как взвешенная сумма, подобно весам, используемым при расчете StatScore. Соответствующие веса, а также число и состав услуг, учитываемых при расчете GlobalNetScore, могут быть выбраны по усмотрению операторов, если рассматривается сценарий внутреннего эталонного тестирования (рисунок 1).

Как вариант, веса можно определить исходя из результатов анализа пользовательской статистики (например, в рамках краудсорсинга), типов используемых услуг и активности пользования ими в зонах различных типов.

Однако глобальная оценка будет корректной только при наличии подробного описания и мотивировки выбора KPI или KQI и их весов в составе каждой услуги, а также веса каждой услуги в глобальной оценке. Более того, как отмечается в основном тексте настоящей Рекомендации, значения StatDiff должны обрабатываться с учетом субъективной значимости различий в тех или иных KPI или KQI для пользователя и устанавливаться равными 0, если, несмотря на возможную статистическую значимость, эти различия не обеспечивают лучших показателей работы сети с точки зрения пользователя.

Библиография

- [b-ITU-T E.802] Рекомендация МСЭ-Т E.802 (2007 г.), *Принципы и методики определения и применения параметров QoS.*
- [b-ITU-T P.10] Рекомендация МСЭ-Т P.10/G.100 (2017 г.), *Словарь по рабочим характеристикам и качеству обслуживания.*
- [b-ITU-T P.863] Recommendation ITU-T P.863 (2018), *Perceptual objective listening quality prediction.*
- [b-ITU-T P.1203] Recommendation ITU-T P.1203 (2017), *Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport.*
- [b-ITU-T P.1401] Recommendation ITU-T P.1401 (2012), *Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models.*
- [b-ETSI TR 102 581] ETSI TR 102 581, V1.2.1 (2015), *Speech processing, transmission and quality aspects (STQ); A study on the minimum additional required attenuation on the antenna path of the field test equipment.*
https://www.etsi.org/deliver/etsi_tr/102500_102599/102581/01.02.01_60/tr_102581v010201p.pdf
- [b-Mehta] Mehta, S. (2014). *Statistics topics.* CreateSpace. 160 pp.
- [b-Shaum] Spiegel, M.R., Schiller, J.J., Srinivasan, R.A. (2013). *Schaum's outlines: Probability and statistics*, 4th edition. New York, NY: McGraw-Hill. 424 pp.

СЕРИИ РЕКОМЕНДАЦИЙ МСЭ-Т

Серия А	Организация работы МСЭ-Т
Серия D	Принципы тарификации и учета и экономические и стратегические вопросы международной электросвязи/ИКТ
Серия E	Общая эксплуатация сети, телефонная служба, функционирование служб и человеческие факторы
Серия F	Нетелефонные службы электросвязи
Серия G	Системы и среда передачи, цифровые системы и сети
Серия H	Аудиовизуальные и мультимедийные системы
Серия I	Цифровая сеть с интеграцией служб
Серия J	Кабельные сети и передача сигналов телевизионных и звуковых программ и других мультимедийных сигналов
Серия K	Защита от помех
Серия L	Окружающая среда и ИКТ, изменение климата, электронные отходы, энергоэффективность; конструкция, прокладка и защита кабелей и других элементов линейно-кабельных сооружений
Серия M	Управление электросвязью, включая СУЭ и техническое обслуживание сетей
Серия N	Техническое обслуживание: международные каналы передачи звуковых и телевизионных программ
Серия O	Требования к измерительной аппаратуре
Серия P	Качество телефонной передачи, телефонные установки, сети местных линий
Серия Q	Коммутация и сигнализация, а также соответствующие измерения и испытания
Серия R	Телеграфная передача
Серия S	Оконечное оборудование для телеграфных служб
Серия T	Оконечное оборудование для телематических служб
Серия U	Телеграфная коммутация
Серия V	Передача данных по телефонной сети
Серия X	Сети передачи данных, взаимосвязь открытых систем и безопасность
Серия Y	Глобальная информационная инфраструктура, аспекты протокола Интернет, сети последующих поколений, интернет вещей и "умные" города
Серия Z	Языки и общие аспекты программного обеспечения для систем электросвязи