IP Video Infrastructure Testing

Test Methodologies

EANTC

August 2009

Evolution of IP Networks: IP Video Services

IP networks were not always the obvious candidates for video transport technologies. As IP based packet services proliferate and infiltrate technologies such as voice, mobile and even storage, video applications are also making the move to the packet world. Service providers are finally comfortable taking full advantage of their IP/MPLS network to converge the variety of their service offerings. IP Video enables service providers to open a new revenue steam not only with IPTV, but also other video related products such as Telepresence and Video on Demand. These new services are likely to be scrutinized both by the end customers, who are used to the quality of the cable TV experience, and by the broadcasters who must adhere to a stringent set of transmission guidelines and service quality.

### EANTC IP Video Testing Experience

In 2007[1] and again in 2009[2] EANTC was contracted by Light Reading online magazine to test Cisco's IP Video infrastructure solutions. In both cases EANTC worked with Spirent Communications to execute the tests and to design realistic scenarios with massive scale. Based on EANTC's IP Video testing experiences in these projects, in addition to other video-based tests, we present this white paper describing the building blocks and the fundamental methodologies behind IP Video testing.

MPEG – The Technology

Today, when speaking about video transmission over IP, we usually refer to MPEG-2 or MPEG-4 compressed video transmitted using the MPEG Transport Stream (MPEG-TS) format (standardize in MPEG-2 Part 1 (ISO/IEC 13818-1) also known as ITU-T Rec. H.222.0). The MPEG container format was invented to transport videos through media with a high bit loss probability and is used in broadcast applications such as Digital Video Broadcasting (DVB) and Advanced Television Standard Committee (ATSC).

The video (either MPEG-2 or MPEG-4, also known as H.264) is encoded and packetized into elementary streams (PES). These elementary streams are multi-

1. http://www.lightreading.com/
   document.asp?doc_id=126173
2. http://www.lightreading.com/
   document.asp?doc_id=177356&

plexed together with audio streams and data to the MPEG transport stream. Each MPEG-TS packet is 188 bytes long and is normally encapsulated into RTP and UDP packets.
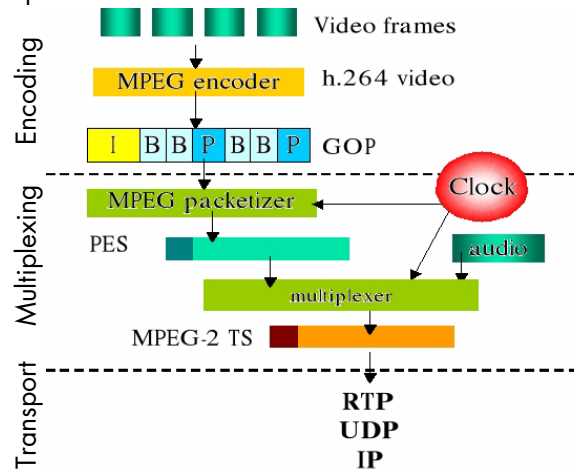
Two different technologies are normally used to deliver IP Video services over packet networks. The choice a service provider has between the two is often dictated by the type of application it is intending to deliver. Broadcast television delivery over IP (normally called IPTV) uses IP Multicast as its delivery mechanism while Video on Demand or other viewer specific services (video surveillance for example) use unicast. These days with growing success of several services we see some service providers moving to a hybrid models – for example near Video on Demand (nVoD) in which VoD content is sent at fixed intervals to a group of viewers. The viewers are then expected to start watching the show as a certain designated time slot. The benefit is that the service provider can use IP Multicast to distribute nVoD hence saving on bandwidth needs.

Getting Started - Test Setup and General Methodology

We tailor EANTC's tests to the specific goals the vendor or service provider has and to the specific system under test (SUT). Nevertheless, IP Video infrastructure tests follow a set of building blocks which apply to every case. The building blocks are described in this white paper. Each section describes why the particular topic is important, along with the test methodology.

Test Parameters

Several parameters serve as the benchmark for IP Video testing. Three parameters specifically influence all IP Video systems and are often discussed in the various specifications:

- Latency – Impacts live broadcasts by delaying viewing

- Packet Delay Variation (PDV) – A high PDV requires the end system (for example a Set Top Box) to store a number of frames in memory (i.e. buffer) hence influencing the design of the end device. The more memory is needed in the system the more expensive it will become.

- Packet Loss – IP Video uses UDP as transport protocol. UDP as no inherent retransmission mechanisms (as TCP has) which means that once packets are lost the end-user quality of experience is noticeably decreased.

The various standards that define these parameters are discussed in the appropriate sections in this document.

Test Equipment

There is a variety of test equipment on the market capable of performing different tests of IP Video networks. In the two public Light Reading tests mentioned earlier, we used Spirent TestCenter. However, any test equipment may be used so long as it supports the necessary features and fits the requirements set by your testing goals. Some of these requirements include:

- Generating and tracking stateless and stateful traffic at line rate

- Integrating both types of traffic streams within a single test

- For some of the tests discussed later in this paper the ability to perform video analysis on live traffic is imperative

- Key metrics such as delay, packet delay variations and loss are an absolute must for test equipment. Preferably, all three parameters should be collected at the same time for correlation since video is dynamic and is hard to recreate in consecutive test runs

Additional requirements for test equipment are the ability to create realistic multi-protocol stacking scenarios (for example HTTP over PPPoE over MPLS which is a typical residential user data stack).

Using Black Box Testing for IP Video

Black box and white box testing should be a familiar term not only for those in the telecommunications industry, but also for those experienced in technology in general. White box indicates that the tests are designed considering the different aspects of "how-it-works" to specifically test areas of interest or potential weakness. Black box tests examine the behavior expected of the SUT without the knowledge of how it works. For example, a white box test of an IP Video infrastructure may include testing aspects of the specific multicast signalling protocol used in the test network. A black box test, on the other hand, would only look into the quality of the IP Video service at the source and at the customer site, disregarding the multicast signalling protocols used.

If one considers the different IP Video transport mechanisms, suggested multicast models, and other best practices that exist for different network components, one will quickly realize that there is a vast array of valid IP Video infrastructure models. The top priority for providers deploying IP Video is of course the experience of the end user. This paper generally describes black box test methodologies which allow IP Video testing from an end user perspective regardless of the underlying technologies of the solution.

Design Verification

Testing a system in telecommunications should be, like testing anything else, a scientific process. The process should be precise, accurate and comprehensive. Narrow scope tests can reveal a lot about a specific function, but when testing something such as a full IP Video network solution, the tests are best done in a realistic scenario by constantly asking the question - how will the provider use this? What is likely to be configured in unison? What will the customer spread look like?

Service Emulation

The performance of a device or system is likely to behave differently when using only one traffic type as opposed to a mix. It is important to set up the testing, the testbed, and to configure the tester such that the conditions emulated during the test will be as close as possible to the intend use of the network or the device. The important questions to ask are:

- What services do I intend to provide with this network?

- What are the frame sizes for each service? Is there a frame size mix that I already know about based on the existing network?

- What is the bit rate for each service per user?

- What types of users do I plan for this network to serve? (e.g. business, residential, broadcasters)

- How many users do I plan to connect to each customer facing port?

- How many IP Video channels?

Testing IP Video Performance

Before looking into video specific application aspects it is important to create a baseline performance confidence of the system under test. More specifically, this means multicast performance testing. Multicast benchmarking is well standardized by the Internet Engineering Task Force (IETF) in a Request for Comments (RFC) numbered 3918 "Methodology for IP Multicast Benchmarking". The tests described in the document borrow from an earlier RFC that is considered a fundamental performance test suite - RFC 2544 "Benchmarking Methodology for Network Interconnect Devices."

Baselining the System

The standard model for IP Video Broadcast distribution uses an IP multicast group for every TV channel. This is exactly what multicast was designed to do - to efficiently distribute packets from a single source to multiple viewers (or subscribers if you will).

The important aspects of the multicast distribution system to verify are:

- The number of groups the system can support. The number of groups is normally directly related to the intended amount of TV channels the system is designed to deliver. Often, for resiliency reasons, two groups are used to distribute the same content. Multicast group count influences two factors in a system: the edge system must be able to support the required number of IGMPs and the network must be able to maintain state on the number of groups through its multicast routing protocol (for example Protocol Independent Multicast - PIM).

- The number of PIM neighbors the system can support. Depending on the network size and design each router must maintain logical connections to the other PIM routers. It is important to verify that these connections can be maintained,

with multicast update messages adhering to the size of the groups supported in the network. Losing PIM neighbors would translate to a large group of customers not being able to receive their video content.

Once the control plane characteristics of the system are known further testing should be performed. This time the performance of the system should be the focus.

Basic Multicast Black Box Test Setup
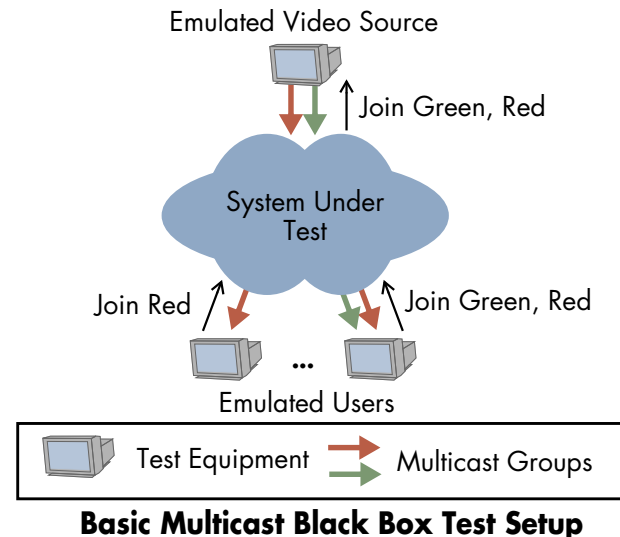
Throughput Performance

Several performance characteristics should be verified for any system designed to support IP Video services: delay, frame delay variations and throughput (zero packet loss state of the system). RFC 3918 does a good job of providing guidelines for these testing.

The first test to perform when taking the thorough route is aggregated multicast throughput. This test is designed to unearth the system's multicast only performance. This enables the tester to collect data points about the system under test and to answer the questions: How does the system behave in an extreme case where only multicast is sent? What is the latency and packet delay variation through the system? Can the system support line rate multicast distribution?

Such a test really puts stress on the last hop router. The router must replicate multicast traffic to all receivers downstream from it. This is by no means an easy task so designing the test to really verify that the underlying architecture meets the requirements is important. For instance, systems often behave differently when multicast replication is performed between line cards than when the same line card is both the source and receiver of the multicast traffic.

Since very few networks are designed to support only multicast, once the above testing have been completed the next step is to verify that both multicast and unicast can exist in tandem in the system. Even in systems that support only IP Video it is likely that Broadcast TV will be transmitted using IP Multicast while VoD will use unicast hence the importance of this test.

RFC 3918 describes a methodology for testing throughput performance of a system using both multicast and unicast traffic. There are some tweaks to certain metrics that can be made to make a more effective video performance test. For example, the RFC recommends several frame sizes starting from 64 bytes up to the Maximum Transmit Unit (MTU). Since video traffic of virtually any type will ideally use large frame sizes, a throughput test using 64 byte frames will simply not give you results that are valuable in this context. The RFC also leaves open the issue of ratio between unicast and multicast traffic. This ratio should be decided based upon the intended network design. In recent testing, based on the intended use of the network, we used the following parameters:

- Unicast to multicast ratio: 3:7 – in triple play service deployment it is often the case that the IP DSLAM receives all IPTV groups hence the ratio is heavily loaded towards multicast

- Frame sizes (Bytes): 1024, 1280, 1518

- Replication factor: 1:120 – this parameter will be influenced by the port capacity of the last hop router in our case 120 egress ports

- Number of multicast groups: 240 – the parameter is derived from the number of channels offered

While the value of the test results will increase as the variables are customized to the intended scenario, the above recommendations can be used as a suggested default. The results that are most interesting for such a test, as any performance test, are frame loss, latency, and frame delay deviation where frame loss should be zero, and latency and frame delay variation expected results depend on the requirements of the video encoders used in the design.

Service Resiliency

It is impossible to eliminate the possibility of a link or node failure in any network. The risk of a power failure, human error, hardware failure, etc, may be reduced, but not completely removed. The solution is to build mechanisms in the network which can quickly recover from such a failure.

When testing for resiliency the result is typically the out-of-service time experienced by the failure. A good resiliency mechanism will introduce a short out-of-service time hopefully within the tolerance levels of the video system. Out of service time is normally calculated according to the amount of frames that were dropped from the test traffic. That number is well documented by the Broadband Forum (in its previous incarnation as the DSL Forum) Technical Recommendation document called TR-126 "Triple-play Services Quality of Experience (QoE) Requirements" which specified acceptable frame loss for IP Video traffic.

### Measuring Out of Service Time

There are two ways to measure out of service time.

1. Calculate using frame loss count: If we assume that during the failure condition no frames are forwarded and also assume that the transmitting bit rate is constant, we can calculate the out of service time by dividing the measured frame loss by the frame rate of the sent traffic stream.

   In some cases there are several drawbacks to using frame loss counter to calculate the out of service time. Imagine that the service does not return to its full rate state immediately, but is ramping up slowly. It is also possible that the service returned to normal operations right away, but still had some short interruption until the system stabilized completely again. In such a case calculating the out of service time using the lost frame count is only a best guess.

2. The second method of measuring out-of-service time can provide a more accurate understanding of the effect of a failure on the service. It involves using an analyzer with high-resolution, real-time counters. It records in very short intervals the received frame rate. With this information it is possible to identify the out of service time with a high degree of precision and also to graph the effect over time. This methodology requires an analyzer that can support high resolution real time counters of frame rate statistics – not a trivial task for analyzer vendors.

The following figure shows two examples of service outages. Picture a) shows the case where calculation based on frame loss would result in an accurate value for the out of service time. Picture b) shows a case where out of service time (OoST) calculation based on frame loss would result in a much shorter OoST compared to the actual service interruption the end user will experience.

$$OoST = \frac{LostFrames}{SentFramerate}$$
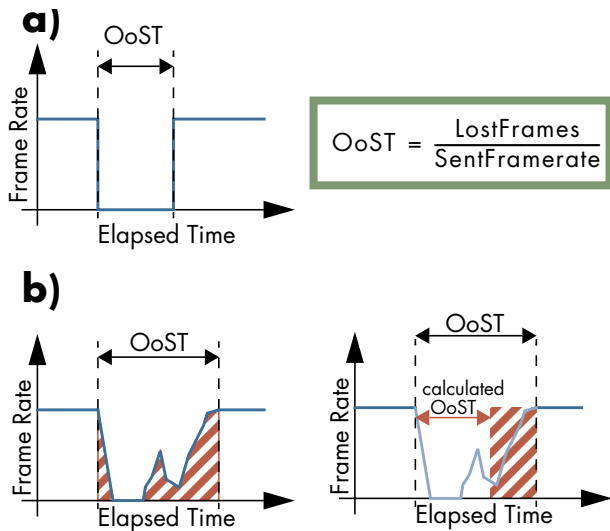
Out of Service Time Calculations

Link and Node Failure

The path between the subscriber and the source (say video headend) consists of links and nodes. Both elements are susceptible to failure. A classic failure situation that carriers worry about is the "Backhoe" scenario - a backhoe digging in the ground and accidently severing a fiber duct and with it the connection between two Points of Presence (PoPs). Node failure refers to the failure of an entire device (router or switch).

When verifying that infrastructure is ready to transport IP Video services it is important to verify that it is not only able to transport the IP Video traffic with adequate quality but also that in a case of a failure the network is able to bypass the failed node or link and continue providing the service to the end users. The Broadband forum TR-126 defines for example, that the average IP Video stream packet loss rate for MPEG-2 at 5 Mbit/s is $5.26E^{-06}$. For MPEG-4 traffic at 3 Mbit/s the same standard sets $5.85E^{-06}$ for the same value. Both parameters are more stringent than Voice over IP frame loss allowance for reasons that are inherit in the design of the MPEG transport streams. In essence since several MPEG packets could be used to fill a single IP packet (the former being 188 bytes in size and the latter easily 1500 bytes), the loss of any such packet could have a devastating effect on the end user Quality of Experience (QoE). A single lost MPEG I-Frame is very likely to result in a visible artifact on a TV screen which is why additional resiliency mechanisms are often used in networking supporting IP Video applications.

Both Link and Node Failure tests must be repeated several times and in the correct order. This means that a test run is comprised of two phases - the failure simulation and the recovery. We recommend repeating this test phase at least three times to gather some confidence level that the results recorded are representative of the system behavior and not an outlier. If the tests produce results that are not within the confidence level described in the test plan it is recommended to repeat the tests and to let the engineers supporting the engagement find the source of the unexpected results.

IP Video Source Redundancy

Having redundant video sources is considered best practice. If the video source experiences a failure condition all IP Video network services will be interrupted and the customers upset. We will consider two models for solving this issue based on our recent testing experiences: providing redundant video sources, or headends, and providing multiple copies of video from a single headend throughout the network. Additional methods exist, however, they were out of scope in our recent testing experiences.

Redundant Headends

One of the benefits of using IP Video for broadcast TV distribution is the multitude of technologies that can be used to offer a level of resiliency to the headends. One such mechanism is using IP anycast.

Briefly stated, anycast is a network addressing and routing scheme in which, using routing protocols, data is routed to the nearest destination. Translated to our discussion here, a network operator can configure identical IP addresses for its video headends letting the routing protocol choose which headend to use as long as more than one is active. Once the active headend fails, the secondary headend is used. The two must be synchronized of course so that the subscriber will not experience the failover.

When testing the effect of a headend failure we can use the same information and methodologies described for the Link and Node failure tests. One important component to remember is that each system will react differently to the failure simulation. If the failure condition is simulated by removing the cable connecting the headend to the first hop router, the router is likely to quickly react and, based on the loss of signal (LOS), to withdraw the prefix from its IGP. But is this a realistic scenario? That question can only be answered based on the Video data center design in your specific test.

Redundant Video Streams

Networks today aim for high resiliency goals as described in this paper. However, since a single frame loss is fatal to a video decoder, it is desirable to bring the traditional 50 milliseconds failover detection and recovery time down to 0 (no loss to the subscribers does not mean no loss in the network, but this will become clear as you read on).

The idea of this infrastructure model is to duplicate a video stream and send two copies into the network. The network should then send each copy through a unique path. The advantage here is that a downstream device can constantly monitor both videos even though only one stream is required. When a stream experience problems due to a failure in the network, the decoder already has a separate unaffected video stream to switch to without waiting for failure recovery technologies. Hence (when this works properly) the end user will experience no adverse effect even though the network might have.

The test procedure is straight forward. Once the infrastructure is configured and video streams are flowing through the network and received by each tester interface, disconnect one of the two interfaces sourcing identical video and measure frame loss and video quality at the receiver ports. Then, repeat the test disconnecting the other link.
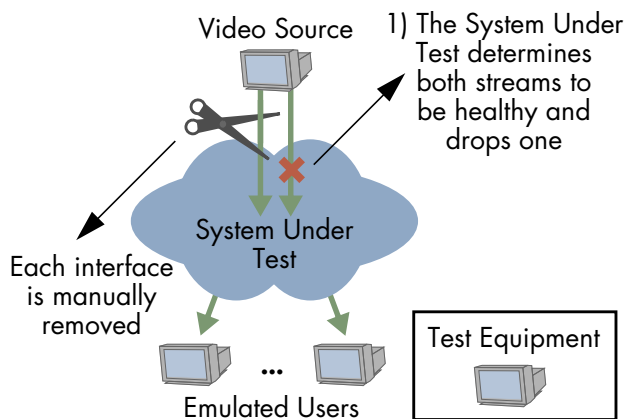
If the system under test does in fact make use of both video flows to truly provide lossless video, both 0 frame loss and no changes in video quality will be observed throughout the test. To note, this procedure assumes that each of the two identical video flows will enter the network on a separate interface. If this is not the case, information relating to how each flow traverses the network must be known in order to remove a link which only affects a single flow.

IP Video Prioritization and Congestion Avoidance

The last, but not least, aspect of IP Video infrastructure testing is the system's behavior in the case of congestion. Lets face it, networks sometimes experience congestion. The cause of such events could be phenomenal success of a service, a distributed denial of service attack (DDoS) or failed capacity planning. We also see very short lived congestion scenarios that are the results of the system's buffering or queueing behavior. Converged networks are likely to use a method to differentiate between packets belonging to one service from the next (i.e. packet markings) and another mechanism to prioritize these packets in the case of congestion (i.e. Quality of Service).

When testing the system's ability to handle congestion scenarios gracefully one must first verify that the system is able to identify packets and mark them as belonging to different classes. The methodology requires careful planning since the tester must generate packets that will match the Device Under Test (DUT) configurations. It is also advised to send traffic for the maximum expected services the DUT should be supporting at the maximum zero frame loss line rate. For example, in triple play deployment scenarios, broadcast video, Voice over IP, VoD, and high speed Internet access, are all potential candidates for their own class of service.

Once the system has been verified for correct marking of traffic the next stage of testing can commence. In this phase you should create congestion scenarios in the system that will require it to prioritize your video traffic. Monitoring two parameters is important: latency and frame loss. In essence the two are solid indicators whether the system is behaving as expected. When latency increases due to congestion, as a tester you can be sure that some frames are spending extra time sitting in a queue somewhere expecting to be processed and forwarded. Depending on the design this might or might not be acceptable.

The second parameter, frame loss, is an indication whether the system is functioning correctly. As we discussed before, losing packets from video streams could directly translate to angry customers. Therefore, when trying to prioritize IP Video traffic it is important that the system will not drop IP Video packets in the case of congestion. When the test results show that IP Video traffic did not experience increased latency or frame loss you can consider the test has passed.

Video Analysis and Monitoring

MPEG Monitoring

Service providers are interested in knowing the impact several network conditions actually have on the quality of the transmitted video. These are important in order to be able to understand and parameterize the end-customer's change in quality of experience.

An essential approach to measure triple play (or now becoming known as 'multiplay') architectures is therefore to emulate network conditions as closely as possible to the real network. This task can be done by using advanced impairment generation devices which are able to simulate most network conditions needed.

TIA-921A and ITU G.1050-2007 describe the parameters reflecting several levels of network conditions such as transfer delay, congestion and queuing. Current impairment generation devices on the market already have implemented predefined profiles reflecting the different scenarios described in ITU G.1050-2007 and TIA-921A.

Accompanying the impairment generators are traffic generators that can generate and monitor video streams. The test device needs advanced analyzing capabilities to assess the video quality. There are some minimum requirements such a device must fulfill.

The European Telecommunications Standards Institute (ETSI) published measurement guidelines for DVB written in the technical report TR101-290. Video monitoring devices should at the very least implement the predefined first and second parameter sets tests defined in the document (such as continuity count errors). In addition, RFC 4445 proposes the Media Delivery Index (MDI), which describes measurement methods to assess delivered video quality without taking the content itself into account. It is only based on network and MPEG-TS header parameters.

Network and transport parameter measurements are not sufficient by themselves to acurely asses the user perceived video quality. This measurement is handled by the Mean Opinion Score (MOS), which provides a numerical indication of the perceived quality and is available for both video and voice. The MOS expresses the perceived quality as a single number in a range from 1 (bad quality, very annoying impairment) to 5 (excellent quality, imperceptible impairment). In principle there are two methods to measure MOS: reference based and non-reference based.

Reference based methods compare the video received at the end with a reference video from the source. The ITU-T published this approach in its recommendation J.247 "Objective perceptual multimedia video quality measurement in the presence of a full reference" for telecommunications services delivered at 4 Mbit/s or less and in J.144 which focused on video quality estimations for television video classes (TV0-TV3), and multimedia video class (MM4).

Here there are several MOS related implementations on the market. They take into account network related parameters (such as packet loss and packet to packet delay variation PPDV - Jitter), content parameters (such as bit rate, I, B and P, Codec type, frame size) as well as MPEG-TS information when calculate MOS. ITU-T SG12 is currently working on standardizing perceptual quality models based on parametric bitstream information (P.nams and P.nbams).

Summary

There are many types of IP Video applications in the market these days: Video conferencing, Video telephony, Video surveillance, Network based Personal Video Recorder, broadcast contribution and video distributions and more. The test methodologies described in this document are by and large universal to all IP Video applications. Nonetheless, each system requires its own unique test plan tailored to its intended use and design. With the building blocks described here your imagination and technical understanding are all that you need to successfully test. You can choose to perform the testing alone, or, contact us, we will be happy to help.

About EANTC

The European Advanced Networking Test Center (EANTC) offers vendor-neutral network test services for manufacturers, service providers and enterprise customers. Primary business areas include interoperability, conformance and performance testing for IP, MPLS, Mobile Backhaul, VoIP, Carrier Ethernet, Triple Play, and IP applications. Testing is offered in our facilities or on site.

EANTC AG: Einsteinufer 17, 10587 Berlin, Germany
info@eantc.com, http://www.eantc.com/