Performance comparision of secure and insecure VoIP environments

Voznak Miroslav
CESNET,z.s.p.o.
Zikova 4
Prague, Czech Republic
miroslav.voznak@vsb.cz

Rozza Alessandro
University of Milan
Via Comelico 39/41
Milan, Italy
rozza@dico.unimi.it

Nappa Antonio
University of Milan
Via Comelico 39/41
Milan, Italy
nappa@security.dico.unimi.it

ABSTRACT
This paper deals with techniques of measuring and assessment of the voice transmitted in IP networks in secure and insecure environment using different virtual testbeds and a real implementation based on OpenSer. We realized our real platform, in order to understand how the voice services in IP network are affected by using the secure IP environment. The real performance test was implemented between VSB-Technical University in Ostrava and University degli studi of Milan.

We have described a secure solution based on Virtual Private Network and how this security can influence the DoS resistance. Our aim was to underline how the voice performances could be influenced by DoS attacks and, how much, the usage of a VPN can decrease the infrastructure resistance while is overloaded by DoS attacks. Following to this, we have pointed out the advantages and the disadvantages of the adopted security measure and we compare the performances of both solutions. Further we have described two virtual testbed, one developed using a traffic emulator and the second one based on a network simulator. Both the virtual environments were implemented in secure and insecure way. The performance evaluation of the VoIP applications is quite complex, for this reason we adopted an application Ixchariot, that is able to compute such complex evaluation indexes, such as MOS and R-Factor.

Keywords
Voice over IP, VoIP security, performance test

1. INTRODUCTION
In the last five years, the growth of voice services over data networks, and especially over the Web, has reached considerable levels. The fast expansion of a technology branch usually raises a number of security issues due to the fact that many of the components of such projects are not generally ready to reach high expansion levels. In this document, we have examined a real VoIP implementation based on Openser, a famous application used by a large number of Voice over IP providers. Our analysis has taken into account characteristic performance indicators of VoIP protocols, in order to evaluate quality of service decreasing due to DoS attacks. We analysed a secure solution based on Virtual Private Network, and the DoS resistance. Furthermore we described two virtual testbeds, one developed using IxChariot and the second based on Ns-2. Ixchariot is a software solution that simulates the VoIP traffic and parameters over a real network environment. Instead Ns-2 is a simulator that can design complex and large-scale network with custom parameters. Both the virtual environments were implemented in a secure and in an insecure way. The performance evaluation of the VoIP applications is described in [1] and [2], we adopted IxChariot that can compute evaluation indexes, such as MOS and R-Factor. On this virtual environment we performed experiments aimed to compare the obtained results with indexes extracted from real tests. In the testbeds using Ns-2 the simulation worked on a single PC and it reproduced our real environment configuration and the desired traffic conditions. It was necessary to use this simulator in order to generalize our results and compare them to large-scale simulations in the future. Virtual Private Network (VPN), assuring confidentiality and strong authentication (using appropriate keys exchange), is able to provide a consistent security defence solution against the first two threats. As well documented in [3]. The security assurance introduced by the VPN does not provide a solution for DoS technique based attacks. We had to consider that the VPN introduces more traffic on the network. For these reasons we wanted to observe if this security solution has some impact on DoS resistance. During our Ns-2 case simulations, we did not have a specific implementation for VoIP application layer. For this reasons during these tests we were able to simulate just generical DoS attacks using a traffic overload to represent the VPN presence. To keep correspondence with attack implemented in Ns-2, the attack simulated in Ixchariot and the real environments was realized by filling bandwidth with UDP packets.

2. TEST ENVIRONMENT
We realized our real platform, in order to understand how the security measures of VoIP such as VPN can reduce the DoS resistance and so bias the speech quality. During the execution of such real-tests we used only a specific DoS attack, focalising our attention on the different performances. The first problem we encountered was about the performance indicator to use.

R-value lower limit	MOS	Speech Transmission Quality Category	User Satisfaction
90	4,34	Best	Very satisfied
80	4,03	High	Satisfied
70	3,6	Medium	Some users dissatisfied
60	3,1	Low	Many users dissatisfied
50	2,58	Poor	Nearly all users dissatisfied

Table 1. The R-factor / MOS Comparison.

The most famous parameters are the MOS “Mean Opinion Score" and the R-Factor. The first one is a 1 to 5 index, where 5 is the best quality speech. This parameter is subjective. On the contrary the R-Factor is a scalar based on the E-Model, a computational model for performance evaluation of data networks. The scale of the R-factor is up to 94 for narrowband and up to 120 for wideband codecs, a conversion between MOS and R-factor is possible. The extraction of these indexes is quite complex and we used a software suite called Ixchariot to make simulations on our platform and compute comparable results. Also, in order to generalize our results, we used Ns-2 with integrated framework for data Collection and Statistical Analysis for testing large bandwidth environment.

2.1 Real Test

The basic configuration of the real tests required 2 PCs provided by GNU/Linux Debian, one was based at the University of Milan, Italy and the other PC at the VSB-Technical University of Ostrava, Czech Republic, the bandwith was limited to 10 Mbit/s by traffic shaping. OpenSer was installed in Ostrava. **OpenSer** is an open source SIP server/router. It can be used on systems with limited resources as well as on Service Provider servers. It is able to manage up to thousand calls per second. In addition to OpenSer, we needed another software to generate Sip/Rtp traffic. We decided to adopt Sipp, an Open Source test tool/traffic generator for the SIP protocol. It includes basic user agent scenarios (UAC and UAS) and establishes and releases multiple calls with the INVITE and BYE methods. It is possible to build custom XML files to describe from very simple to complex call flows. The results of a test are dynamically displayed during the execution (call rate, round trip delay, and message statistics). One of the more important feature is the possibility to send media (RTP) traffic through RTP echo and RTP/pcap replay. Media can be audio or audio+video. We used this option to generate real calls with G.729 and G.711Alaw audio codecs. The audio codecs used, G.729 and G.711Alaw, are two different formats of audio compression. The first one, is commonly used with VoIP devices, the standard configuration of G.729 operates at 8 kbit/s bitrate, a low bandwidth requirement. The G.711Alaw codec offer a higher level of quality respect of the G.729, but takes more bandwidth. The level of bitrate of this codec is 64 kbit/s. This codec is a standard audio compression format defined by the ITU-T. The difference of voice quality evaluation between the codecs is stated as 10% [4]. The choice of two different codecs was made to underline that a low bandwidth requirement codec is of course offering less voice quality, but offers more DoS resistance. To obtain valid G.711Alaw and G.729 traffic, it was necessary to capture one of our test call with Wireshark. The captured traffic was used during the tests to generate the RTP traffic. The test configuration was deployed to realize 25, 50 and 100 concurrent calls with the duration of 30sec. Every concurrency step (25,50,100) was repeated one-hundred time for both codecs (G.711 and G.729). For every codec the test has done with or without VPN. We adopted **OpenVpn**, a simple security solution, for protecting the SIP signalling and the RTP data flow using the TLS protocol. Using both the secure or the insecure configuration, the testbed network has been overloaded with 4 Mbit/s of UDP traffic using **Iperf**. Iperf reports bandwidth, delay jitter, datagram loss. This software permits to generate a pre-defined amount of traffic for a pre-defined duration. This last feature was very important for us, in order to configure the tests over stress conditions. During the test we tried to attack the network in a specific way, mentioned in Section 3, to understand the level of resistance of the network itself. As said before, the OpenSer server was based in Ostrava and the Sipp client in Milan. On the machine in Ostrava we also configured another instance of Sipp, in order to answer to the calls addressed by OpenSer. Both universities are connected to the national research and education networks which are linked by multi-gigabit pan-European communication network, called Geant2. It is a significant advantage to be part of high speed network because the end-to-end delay between our endpoints has been approximately 30 ms.

2.2 Testbed using IxChariot

IxChariot is a software, produced by Ixia, useful to predict device and system performance under realistic network loading conditions. The security solution adopted was OpenVpn. The test environment was built with the IxChariot console and two IxChariot endpoints (Installable under several operating systems). The IxChariot console allows selecting several test configurations for IPv4 with and without QoS. At the end of the test, for each possible configuration, it is possible to obtain measurement of the throughput, jitter, MOS and R-Factor. The concurrent calls were repeated one-hundred times for both codecs (G.711 and G.729). We have noticed that the best way to execute a test is by the batch procedure, because in this way the final results was sent to the console only at the end of the test. In such way it was possible to avoid some influence, due to the result data, during the test [5].

2.3 Simulation with Ns-2

It was necessary to use Ns-2 to generalize our real testbed environment. We replicated the Milan - Ostrava scenario with the same end-to-end delay and with the same bandwidth capacity, in order to evaluate and to compare performance losses under DoS attack. Ns-2 is not provided with an appropriate tool for value measuring indexes such as Packet loss, throughput, IP delay variation, etc. For this reason we chose the Ns-2 Measure module, an integrated framework for data collection and statistical analysis within Ns-2. The Ns-2 testbed was deployed with the same traffic loading of the real testbed, using a CBR (Constant bit rate) application over UDP. In order to simulate the attacks and the security measures we simply increased the value of the CBR in such a way to overload the simulated network and observe the packet-loss and the throughput. Our aim was to obtain a valid Ns-2 model to replicate the VoIP application behaviour, in order to study large-scale attacks and countermeasures in the future. The CBR model offers the advantage to insert in the simulated network the desired amount of traffic. Following this, it was possible to examine the Ns-2 test configuration. Our choice was to simulate the test configuration with 100 concurrent calls. During the secure tests the VPN was computed by incrementing the CBR value of the 3%. This value was decided analysing the data reported in [5], [6] where the results obtained showed that the VPN loading is contained between the 1% and 5% of the RTP traffic total amount.

3. ADOPTED ATTACK TECHNIQUE

In this section we explain in details the specific implementation of our attack in real tests, using IxChariot and Ns-2, and give a summary of our results.

3.1 Implementation of Udp flooding attack

With the use of Iperf it was possible to generate TCP/UDP traffic and address it to any destination as “IP.ADDR:PORT". Our attack

was developed to overload the network with 4 Mbit/s of traffic at each step of the tests both in real testbed as in IxChariot case. The tests with the G.711 codec were the most influenced by the flooding attack because the codec used had a high bandwidth loading. In the worst case (100 concurrent calls + VPN + 4Mbit UDP flooding) this test was repeated 100 times and the average packet loss equalled 15%. In the same case (100 concurrent calls + VPN + 4Mbit UDP flooding), but using the G.729 codec, we registered a 5% packet loss.

3.2 Implementation of flooding attack in Ns-2

To simulate the behaviour of our network using Ns-2 it was necessary to calculate the correct value of CBR. For a call with G.711 codec the bitrate on the application layer is 68.8 kbps with common payloads 160 Bytes. At this point we added the VPN overhead that we stated before as the 3% of the generated traffic.

4. ANALYSIS OF RESULTS

The deployed test in real case, using IxChariot and Ns-2 were realized under attack condition in order to evaluate the performance differences between an infrastructure with or without security measures. The adoption of a VPN and the repeated DoS attacks are of course increasing the packet loss during the transmission and for this reason PLC algorithms are used. The Packet loss concealment (PLC) is a technique used to reduce the effects of packet loss in Voip quality. The PLC technique is different upon the used codec. A simple method used by waveform codecs such as G.711, is to replay the last received sample with increasing attenuation for each replay. The packet loss can influence the I_{E-EF} factor (Impairment Factor) which is part of the R-factor as you can see in the relation (1).

$$R = R_0 - I_S - I_D - I_{E-EF} + A \quad (1)$$

The maximum value of R-factor for narrowband codecs is 94, the overall quality (R-factor) is calculated by estimating the signal to noise ratio of a connection (R_0) and subtracting the network impairments (I_S, I_D , I_{E-EF}) and by adding Advantage factor A. R_0 is derived from original SNR (Signal to noise ratio), it considers non-optimum sidetone, quantizing distortion, overall loudness and other impairments which occur more or less simultaneously with the voice transmission. The delay impairments are included in the parameter I_D as a mathematical summary of transmission delay, talker echo and sidetone. I_{E-EF} is an equipment impairment that considers the influence of used codecs and impairments due to packet loss and rejection. In the case of G.711 without PLC or with PLC bursty packet loss it is possible to observe an issue. The impairment of speech quality appears in specified ranges of traffic between 80 and 90 percentage points. In our case it is very important to notice that the observed value of packet loss changes significantly in the test with VPN, overload traffic and 100 VoIP pairs emulation, because the whole traffic approaches the limit of the available bandwidth. In this case the speech quality depends especially on the packet loss. The particular values of I_E factor are stated in recommendation ITU-T G.113 and each codec is assigned a degradation value. If G.711 is used then $I_E = 0$ while in case of G.729 the value $I_E = 10$, but there are also known relations between the packet loss and I_E factor which are represented in figure 1. The plot of I_E values vs. packet loss in figure 1, shows the I_E factor increasing proportionally to packet loss. The curve for G.711 without PLC (Packet Loss Concealment) indicates a higher value of the I_E factor. As the packet loss increases from 0 to 5%, the I_E value increases from 0 to 55 in case G.711 without PLC and from 0 to 15 in case G.711 with PLC. It shows the effectiveness of the PLC algorithms and how the usage of them can increase the R-factor and consequently the Voice Quality. The PLC algorithms are furtherly subdivided into random and bursty packet loss conditions and they are more efficient in the first case [7]. The third curve in figure 1 describes G.711 with PLC algorithm.

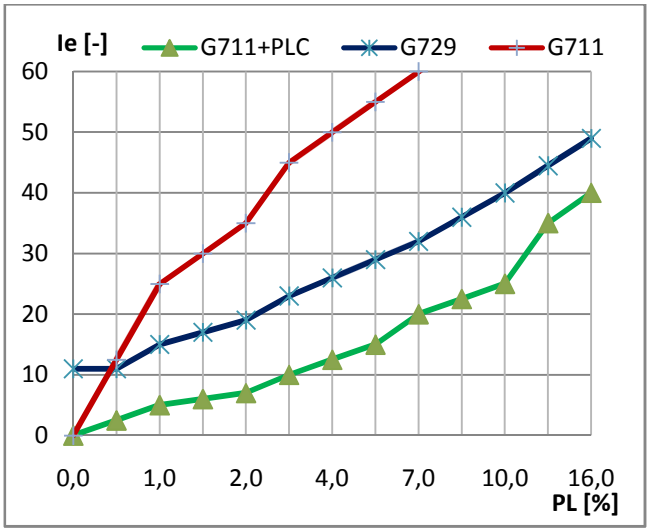

Figure 1. Ie factor impairment due to packet loss

Packet loss can be mitigated by PLC. From the point of view of the Voice quality, a packet loss rate of 2% or lower is not noticeable by the talkers. During our experiments we observed a packet loss higher than 2% in the tests with security measures and traffic overloading. For this reason we investigated how the voice quality could be biased during DoS attacks and the usage of VPN. IP network packet loss distribution can be modelled using a Markov process. A multi-state Markov Model is used to measure the distribution of lost or discarded packets or frames, and to divide the call into “bursts” and “gaps”. The call quality is calculated separately in each state and then combined using a perceptual model, such as in VQmon [8]. The mentioned VQmon does incorporate G.107 compliant implementation of the E-Model. However, we applied a very simple method described in the last revision of G.107 from 2005 [1]. The impairment factor values for codec operation under packet-loss have formerly been treated using tabulated, packet-loss dependent I_E-values (Figure 4). In the last revision, the Packet-loss Robustness Factor B_{pl} is defined as codec-specific value. B_{pl} can be described as the robustness of the codec to packet-loss. Both values are listed in Appendix I of ITU-T G.113 and are available for several codecs. If we consider the Packet-loss Probability as P_{pl}, the I_{E-EF} factor can be calculated using the formula:

$$I_{E-EF} = I_E + (95 - I_E) \cdot \frac{P_{pl}}{\frac{P_{pl}}{BurstR} + B_{pl}} \quad (2)$$

$BurstR$ is the so-called Burst Ratio, when packet loss is random $BurstR$ = 1 and when packet loss is bursty $BurstR$ > 1. For packet loss distributions corresponding to a 2-state Markov model with transition probabilities p between a "found" and a "loss" state, and q between the "loss" and the "found" state, the Burst Ratio can be calculated as:

$$BurstR = \frac{1}{p+q} \quad (3)$$

In our case we assumed a random packet loss distribution and we calculated I_{E-EF} with the values obtained from the real Sipp tests. The achieved results are listed in table below. The results are valid for G.729 a G.711 codecs and the loss in an amount 5% and 15% in the worst-case (VPN+traffic 4Mbps).

valid for 100 concurrent calls	**only VoIP traffic w/o VPN**	**with VPN**	**with traffic 4Mbps**	**with VPN + 4Mbps**
Ie-eff (G729+PLC)	10,4	10,5	10,7	27,7
Ie-eff (G711+PLC)	0	2,2	30,7	35,5
R-factor (G.729+PLC)	82,6	82,5	82,3	65,3
R-factor (G711+PLC)	93	90,8	62,3	57,5

Table 2. Calculated Ie-eff and R-factor.

Once the I_{E-EF} factor was calculated it was not difficult to determine R-factor as an output of E-Model (formula 1) using implicit values of recommendation ITU-T G.107 which are $R_0 = 94{,}7688$, $I_S = 1{,}4136$, $A = 0$, hence we could modify formula 1,

$$R = 93{,}3553 - I_D - I_{E-EF} \quad (4)$$

The model used to estimate I_D is described in [7]. Where it is explained that the effects of delay are well known and easily modelled. Delays of less than 175ms have a small effect on conversational difficulty, then $I_D = 4 \cdot T$, where T is the delay in ms (In our case 30msec). The final achieved values of R-factor correspond to Ixchariot results with an aberration less than 5% and are listed in Table 2. As we can observe from this table the R-factor obtained from our tests shows how the security measures adopted influence the Voice quality under DoS attacks.

5. CONCLUSIONS AND FUTURE WORKS

The real-time applications are very sensitive to packet loss, and each variation occurring on the network can modify and influence the final result of a real-time data transmission, such as a VoIP Call. It is easy to understand that Denial of Service attacks are one of the major security problems in VoIP applications and one of the possible barrier that prevents many businesses from employing this technology. Indeed, as we have seen during our real environment implementation, it is quite simple to apply eavesdrop defence techniques using cryptography but this countermeasure reduces the tolerance to the DoS attacks and the performances of VoIP services. The growth up of the botnet phenomenon - the ability of infected computers networks, called zombies, to perform centralized DDos (Distributed Denial of Service) attacks against determined targets - increases the danger of this threat. With our Ns-2 model, which is correlated to real environment and test environment, it is possible to obtain an approximate amount of the whole IP traffic including VoIP. This will help us to study the performance behaviour in large-scale attacks and the possible countermeasure effectiveness. It's easy to understand that possible related works in this field could be connected with emerging threats of VoIP Botnets, studying possible countermeasures and analysing how they influence the performance indexes. The study presented in this paper is an extension of a previous work on the impact of security on the quality of VoIP calls [5], [6], [9] and provides a framework to analyse the quality of VoIP calls. The paper presents several new contributions. Firstly, it shows how to calculate R-factor from the packet loss, it brings a fast and easy method for the voice quality assessment (relations 2, 3 and 4). Secondly, it shows how VPN with TLS influences the voice quality.

The executed measurements prove the obvious impact of some secure solutions on the voice quality. The impairment of speech quality appears in specified ranges of traffic comprised between 80-90%. While using the G.711 codec in the insecure environment we observed the shift of R-factor from range "Best" quality to "Low" quality and in the secure environment to "Poor" quality. We did not register the change of quality for G.729 calls in the insecure environment but we observed the shift of R-factor from range "High" quality to "Low" quality using the VPN environment. The threshold values of R-factor are defined in Table 1. We suppose it depends on the size of the block cipher encryption and next investigation could be a challenge for our future research.

6. REFERENCES

[1] ITU Recommendation G.107. E-model, a computational model for use in transmission planning. 2005.

[2] ITU Recommendation P.800.1. Mean Opinion Score.

[3] Porter,T. Practical VoIP Security. Syngress, 2006.

[4] Davidson,J.-Peters,J. Voice over IP Fundamentals. Cisco Press, 2000.

[5] Voznak,M-Nappa,A. Performance evaluation of voip infrastructure. FreeVoice, November 2007.

[6] Bruschi,D. Voice over ipsec: Analysis and solutions. Proceedings of the 18th Annual Computer Security Applications Conference, December 2002.

[7] Clark,A. Modelling the Effects of Burst Packet Loss and Regency on Subjective Voice Quality, 2001,

[8] Clark,A. Extension to the E-Model to incorporate the effects of time varying packet loss and recency. ETSI TIPHON committee, TS 101 329-5 Annex E, July 2001.

[9] Nappa,A-Bruschi,D-Rozza,A.-Voznak,M. Analysis and implementation of secure and unsecure Voice over IP environment and performance comparison using OpenSER. Technical report, Università degli studi di Milano, 2007.