

FACULDADE DE TECNOLOGIA SENAI FLORIANÓPOLIS

CURSO SUPERIOR DE TECNOLOGIA EM REDES DE COMPUTADORES



Análise da qualidade de chamadas VoIP através de reconhecimento de fala

RAFAEL LUIZ KLOCK

Florianópolis (SC)
2018
RAFAEL LUIZ KLOCK

Estudo da qualidade de chamadas VoIP baseado em reconhecimento de fala

Trabalho de Conclusão de Curso apresentado a
Faculdade de Tecnologia SENAI FLORIANÓPOLIS
como requisito parcial para obtenção do título de
Tecnólogo em Redes de Computadores.
Professor Orientador: Felipe Machado Von Hertwig

Florianópolis (SC)
2018
RAFAEL LUIZ KLOCK

...

Trabalho de Conclusão de Curso considerado adequado para a obtenção do título de Tecnólogo em Redes de Computadores, aprovado pela Faculdade de Tecnologia SENAI FLORIANÓPOLIS

20 de agosto de 2018
Florianópolis/SC

Orientador Felipe Machado Von Hertwig
Membro

Coordenador Mauro Cesar Matias
Membro

Professor
Membro

Dedico este trabalho à minha amada esposa.

AGRADECIMENTOS

Agradeço a Deus e a todos que me ajudaram direta ou indiretamente para a conclusão deste trabalho.

KLOCK, Rafael Luiz. **Estudo de qualidade de chamadas VoIP baseado em reconhecimento de fala**. Florianópolis, 2018. Trabalho de Conclusão de Curso - Curso Superior de Tecnologia em Redes de Computadores. Faculdade de Tecnologia do SENAI/Florianópolis, Florianópolis, 2018.

RESUMO

Cada vez mais a tecnologia VoIP está presente em nosso meio através de softwares como Skype, Google, Facebook e WhatsApp por exemplo. Já é utilizada por grandes e pequenas empresas com objetivo de facilitar a comunicação, interligar suas filiais, e também reduzir custos de telefonia. Para que uma chamada VoIP tenha uma boa qualidade, é necessário ter uma boa infraestrutura, processamento adequado, qualidade no ambiente de rede, além de priorização de tráfego de voz através de QoS. Contudo, percebe-se a necessidade de monitorar e avaliar constantemente a rede de fornecedores de serviços VoIP que por sua vez também passam por instabilidades e problemas em sua estrutura interna, sistemas e até mesmo em suas interconexões com outras operadoras. Este trabalho busca criar um mecanismo que visa a avaliação de qualidade de voz de fornecedores e operadoras de serviços VoIP. Tal processo ocorrerá através de um sistema de reconhecimento de fala, o qual possibilitará a análise dos indicadores de qualidade, resultando em praticidade e automatização. Desta forma, será feito o envio de um áudio “frase chave”, que será transmitido através de uma rede A ou B, assim ele poderá ser reconhecido, transcrito e comparado em um servidor remoto que possuirá um sistema de reconhecimento de fala, integrado com PBX IP Asterisk. Com este mecanismo espera-se analisar e identificar se a qualidade do áudio foi afetada por determinada rede ou fornecedor.

Palavras-chave: Voip. Reconhecimento de fala. Qualidade de voz. Asterisk.

LISTA DE ILUSTRAÇÕES

LISTA DE QUADROS E TABELAS

[illegible]

LISTA DE ABREVIATURAS E SIGLAS

ABNT – Associação Brasileira de Normas Técnicas

TCC – Trabalho de Conclusão de Curso

SENAI – Serviço Nacional de Aprendizagem Industrial

SUMÁRIO

| | |
|--------------------------------------|-----------|
| 1 INTRODUÇÃO | 19 |
| 1.1 JUSTIFICATIVA | 20 |
| 1.2 OBJETIVOS | 21 |
| 1.2.1 Objetivo Geral | 21 |
| 1.2.2 Objetivos Específicos | 21 |
| 2 FUNDAMENTAÇÃO TEÓRICA | 22 |
| 2.1 Redes de computadores | 22 |
| 2.1.1 ISO e o Modelo OSI | 22 |
| 2.1.2 Protocolo IP | 25 |
| 2.1.3 UDP | 26 |
| 2.2 VoIP (Voz sobre ip) | 26 |
| 2.2.1 SIP | 28 |
| 2.2.2 RTP | 29 |
| 2.2.3 Codificadores | 30 |
| 2.2.4 Qualidade VoIP | 31 |
| 2.2.2.1 Latência (delay) | 31 |
| 2.2.2.2 Jitter | 32 |
| 2.2.2.3 Perda de pacotes | 33 |
| 2.2.5 Asterisk | 33 |
| 2.3 Reconhecimento de fala | 34 |
| 2.3.1 Computação em Nuvem | 34 |
| 2.3.2 Google Cloud Speech Recog API | 35 |
| 3 PROCEDIMENTOS METODOLÓGICOS | 36 |
| 3.1 Cenário proposto | 38 |
| 3.2 Softwares utilizados | 39 |
| 3.2.1 Amazon AWS | 39 |
| 3.2.2 Instância Linux Ubuntu Server | 40 |
| 3.2.3 Asterisk | 40 |
| 3.2.4 DID | 40 |
| 3.2.5 Shell script | 40 |
| 4 CENÁRIO DESENVOLVIDO | 41 |
| 4.1 AWS EC2 | 41 |
| 4.2 Asterisk | 42 |
| 4.3 Google Speech Recognition | 43 |
| 4.4 Contextos e Extensões | 43 |
| 4.5 Shell-script “gencall_X.sh” | 45 |

| | |
|--|-----------|
| 4.6 Shell-script “analyse.sh” | 46 |
| 4.7 Os áudios | 48 |
| 5 RESULTADOS | 49 |
| 5.1 Áudio 1 | 49 |
| 5.2 Áudio 2 | 51 |
| 6 CONSIDERAÇÕES FINAIS | 54 |
| 6.1 RECOMENDAÇÕES PARA TRABALHOS FUTUROS | 54 |
| REFERÊNCIAS | 55 |
| APÊNDICES | 57 |
| ANEXOS | 58 |

1 INTRODUÇÃO

O setor de telecomunicações detém um papel importante na sociedade e no desenvolvimento do país. Tal dimensão é observável, considerando a facilidade da comunicação no dia a dia das pessoas, e sobretudo no ramo corporativo. Através da Internet, vieram muitas outras facilidades que não se imaginava na sua criação, como por exemplo utilizar a telefonia por meio da rede mundial de computadores. Brito (2013 p,22) conceitua que a Internet “É um dos maiores projetos, senão o maior, já construído pela engenharia humana”. Uma prova disso é que essa tecnologia ainda hoje continua crescendo em ritmo acelerado. É através dela que podemos nos comunicar com o outro lado do mundo com muita praticidade e velocidade. É também um ótimo meio de informação, ela já faz parte de nossas vidas e as vezes nem nos damos conta disso. Como afirma Brito (2013 p.23), a Internet atingiu 50 milhões de usuários em menos de cinco anos desde sua invenção. Podemos perceber que ainda nos dias de hoje está crescendo cada vez mais.

Em paralelo, o uso da tecnologia que trafega voz sobre IP denominada VoIP, também se inseriu no mercado muito rapidamente, e tem crescido a cada ano. Em uma matéria publicada pela TW Solutions, mostra-se que o mercado VoIP está em constante expansão, semelhantemente, em pesquisas realizadas pela empresa Infonetics Research, que é especialista na área de telecomunicações, o mercado VoIP cresce cerca de 8% ao ano. COPPOLA (2016), afirma que “O crescimento da indústria VoIP tem sido fenomenal ao longo dos últimos cinco anos em praticamente todos os países industrializados. Esses dados e informações refletem o aparecimento de aplicativos e softwares que utilizam essa tecnologia às quais estamos tão acostumados e dependentes como Skype, WhatsApp, Facebook Messenger, Twilio, entre outros.

Contudo, existem alguns desafios a serem ultrapassados para que o VoIP possa funcionar com total qualidade e desempenho. Podemos citar aqui, os cuidados com a Infraestrutura de equipamentos de rede, a qualidade de um provedor Voip, a largura e qualidade de banda contratada com as operadoras, e capacidade de processamento de equipamentos envolvidos no tráfego de voz. Para

empresas que dependem unicamente de provedores VoIP como por exemplo Call Centers, ou ainda empresas que fazem a revenda de minutos VoIP, é extremamente importante manter a qualidade de suas chamadas. Deste modo, torna-se relevante a criação de mecanismos que possam medir a qualidade das chamadas, tanto em redes locais, como em redes remotas. O presente trabalho pretende o desenvolvimento e análise de um mecanismo que visa a avaliação da qualidade de ligações, às quais são transportadas por redes de longa distância de provedores Voip. Será desenvolvido como instrumento de medição de qualidade, um sistema remoto de reconhecimento de fala, utilizando-se de um PBX IP Asterisk integrado com um recurso denominado “Google Speech recog API”.

1.1 JUSTIFICATIVA

Com aproximação tecnológica, e o crescimento acelerado do uso de Voip, percebe-se a necessidade de criar maneiras de monitorar e medir a qualidade das chamadas. Na compreensão de Gonçalves (2005, p.3), “a telefonia IP quando atingir massa crítica fará com que o PABX de qualquer empresa possa falar diretamente com o PABX de qualquer outra através da Internet”. São inúmeros os benefícios que a tecnologia de Voz sobre IP pode proporcionar, porém para que funcione de maneira adequada, sem falhas e interrupções, são necessários planejamento, investimentos em infra-estrutura, largura de banda, assim como priorização do transporte dos pacotes de voz através de QoS. Além desses pontos, também nota-se a importância de avaliar a qualidade das redes de computadores as quais não se tem controle, como por exemplo a rede de um fornecedor de telefonia VoIP. Deste modo, justifica-se desenvolver um mecanismo de análise de áudio de voz através de reconhecimento de fala para que se possa ter alguns parâmetros de avaliação e consequente tomada de decisão.

1.2 OBJETIVOS

Serão apresentados a seguir os objetivos gerais e os específicos.

1.2.1 Objetivo Geral

Desenvolver um mecanismo de comparação que seja capaz de analisar a qualidade de voz que é trafegada por determinadas redes IP, sendo elas de fornecedores ou não, através de um software de reconhecimento de fala integrado com Asterisk.

1.2.2 Objetivos Específicos

- Instalar e configurar um ambiente de teste com Asterisk.
- Integrar a API (Google Speech Recog) com o software Asterisk.
- Realizar testes com áudios de diferentes tipos, sendo eles com picotes, falhas, mudos e áudios perfeitos.
- Relatar e apresentar os resultados dos testes.

2 FUNDAMENTAÇÃO TEÓRICA

O intuito deste trabalho é analisar especificamente às questões que envolvem a qualidade de voz em redes de computadores e a análise de áudio através da tecnologia de reconhecimento de fala. Nesta sessão, o trabalho tem como finalidade dar embasamento teórico à análise que será realizada. Desta forma, serão descritos a seguir alguns conceitos em relação a redes de computadores, o modelo OSI, e protocolo IP. Na sequência também serão apresentados fundamentos sobre reconhecimento de fala, Google Speech API, algumas características em relação ao protocolo SIP, Asterisk e VoIP.

2.1 REDES DE COMPUTADORES

Parafraseando Moraes (2003), redes de computadores são dois ou mais computadores conectados entre si compartilhando seus recursos que podem ser desde informações, periféricos conectados a cada um, ou até mesmo recursos de processamento. Pode-se entender então que rede é interligação entre dois pontos possibilitando a comunicação e a troca de dados entre computadores.

2.1.1 ISO e o Modelo OSI

A história da ISO começou em 1946, quando delegados de 25 países se reuniram no Instituto de Engenheiros Civis em Londres e decidiram criar uma nova organização internacional para facilitar a coordenação internacional e a unificação de padrões industriais. Em 23 de fevereiro de 1947, a nova organização ISO, iniciou oficialmente as operações. (ISO, 2018).

“A ISO é a maior organização de padronização do mundo. Ela é responsável pelo desenvolvimento de padrões em várias áreas do conhecimento. Essa entidade é formada por diferentes organizações do mundo todo, sendo que a representante americana é a ANSI (American national Standards Organization). A ISO esteve envolvida de padronização de um grupo de protocolos usado para redes de computadores.” (MORAES, 2003, p. 35).

Ela foi a responsável pela criação de um modelo referência conhecido como “Open Systems Interconnection” ou Modelo OSI. Segundo Moraes (2003, p.36), “O modelo de referência OSI define camadas funcionais que podem ser incorporadas aos sistemas de comunicação que se dizem ‘abertos’.”

“Esse modelo é baseado em camadas. Cada camada possui seu papel dentro dos procedimentos e normas criados pelo modelo OSI; isto inclui protocolos e serviços prestados às camadas adjacentes. [...] Sendo assim, cada camada superior faz uso dos recursos disponibilizados pela camada inferior. O Modelo OSI define, portanto, uma estrutura com sete camadas hierárquicas: Aplicação, Apresentação, Sessão, Transporte, Rede, Enlace e Física” (MORAES, 2003, p. 35).

Figura 1 - Camadas do Modelo OSI



Fonte: Kolb, (2016)

Em seguida serão apresentadas as camadas do modelo OSI sob a visão de Tanenbaum (2003). São elas:

- A. **Camada 1 - Física:** A camada física trata da transmissão de bits brutos por um canal de comunicação. O projeto deve garantir que, quando um lado enviar um bit 1, o outro lado receberá como um bit 1. Nessa camada, as questões de projeto lidam em grande parte com interfaces mecânicas, elétricas e de sincronização, e com o meio físico de transmissão que se situa

abaixo da camada física. Exemplo: Cabos e Fios (Ex.: Cabo UTP / Coaxial, Fibra Óptica);

- B. **Camada 2 - Enlace de Dados:** A principal tarefa da camada de enlace de dados é transformar um canal de transmissão bruta em uma linha que pareça livre de erros de transmissão não detectados para a camada de rede. Para executar essa tarefa, a camada de enlace de dados faz com que o transmissor divida os dados de entrada em quadros de dados (que, em geral, têm algumas centenas ou alguns milhares de bytes), e transmita os quadros seqüencialmente. Se o serviço for confiável, o receptor confirmará a recepção correta de cada quadro, enviando de volta um quadro de confirmação. Exemplos: Switches Comuns (comutadores). Há alguns switches, capazes de funcionar também na Camada 3; Placas de rede.
- C. **Camada 3 - Rede:** A camada de rede controla a operação da sub-rede. Uma questão fundamental do projeto é determinar a maneira como os pacotes são roteados da origem até o destino. As rotas podem se basear em tabelas estáticas, "amarradas" à rede e raramente alteradas. Elas também podem ser determinadas no início de cada conversação; por exemplo, uma sessão de terminal (como um logon em uma máquina remota). Por fim, elas podem ser altamente dinâmicas, sendo determinadas para cada pacote, com o objetivo de refletir a carga atual da rede. Exemplos: Routers (roteadores);
- D. **Camada 4 - Transporte:** A função básica da camada de transporte é aceitar dados da camada acima dela, dividi-los em unidades menores caso necessário, repassar essas unidades à camada de rede e assegurar que todos os fragmentos chegarão corretamente à outra extremidade. Além do mais, tudo isso deve ser feito com eficiência e de forma que as camadas superiores fiquem isoladas das inevitáveis mudanças na tecnologia de hardware. O protocolo de transporte pode operar em 2 modos, Orientados à Conexão ou Não. Exemplos: Computadores, equipamentos para firewall, DNS, roteadores mais poderosos.

- E. **Camada 5 - Sessão:** A camada de sessão permite que os usuários de diferentes máquinas estabeleçam sessões entre eles. Uma sessão oferece diversos serviços, inclusive o controle de diálogo, o gerenciamento de token e a sincronização.
- F. **Camada 6 - Apresentação:** Enquanto as camadas mais baixas se preocupam com a movimentação de bits, a camada de apresentação está relacionada à sintaxe e à semântica das informações transmitidas. Para tornar possível a comunicação entre computadores com diferentes representações de dados, as estruturas de dados a serem intercambiadas podem ser definidas de maneira abstrata, juntamente com uma codificação padrão que será usada durante a conexão. A camada de apresentação gerencia essas estruturas de dados abstratas e permite a definição e o intercâmbio de dados de nível mais alto (por exemplo, registros bancários).
- G. **Camada 7 - Aplicação:** A camada de aplicação contém uma série de protocolos comumente necessários para os usuários. Um protocolo de aplicação amplamente utilizado é o HTTP (HyperText Transfer Protocol), outros protocolos de aplicação são usados para transferência de arquivos, correio eletrônico e transmissão de notícias pela rede. Exemplos de aplicações: Mozilla Firefox, Google Chrome), Thunderbird e Outlook Express.

2.1.2 Protocolo IP

O protocolo de internet mais conhecido como IP (Internet Protocol), opera na camada de rede do modelo OSI, e é projetado para uso em sistemas interconectados de redes de comunicação por comutação de pacotes. O protocolo da Internet prevê transmissão de blocos de dados chamados datagramas de uma origem para um destino, onde eles são identificados por endereços de comprimento fixo (RFC 791, 1981). Podemos notar que o protocolo IP, é fundamental para o fluxo

e transporte de dados tanto em uma rede de computadores local, quanto na Internet, que é a rede mundial de computadores.

“O protocolo IP possibilita o roteamento dos pacotes para que possam transitar pela rede e alcançar determinado destino – ele tem o comportamento de policial rodoviário, controlando o fluxo dos dados da grande rede. O Protocolo IP é responsável pela comunicação entre máquinas em uma estrutura de rede TCP/IP.” (MORAIS, LIMA, FRANCO, 2012).

Na visão de Moraes (2003, p.122), “O protocolo IP tem dominado completamente o mercado mundial das redes de computadores, principalmente por ser amplamente disseminado nas plataformas da maioria dos fabricantes.” Tendo como base o protocolo IP, o UDP é utilizado para transportar dados em tempo real. Veremos a seguir mais detalhes sobre esse protocolo.

2.1.3 UDP

O UDP (User Datagram Protocol), é o protocolo mais utilizado para transporte de tráfego de voz, pois possui uma estrutura simples de cabeçalho facilitando a comunicação em tempo real. Na concepção de Moraes et al (2003), o UDP não possui mecanismos de reconhecimento, e por esse motivo, é extremamente rápido e eficiente.

“Ao contrário do TCP, o UDP é um protocolo não orientado a conexão, pois não contém mecanismos de garantia de entrega das informações. Neste caso, as aplicações é que devem criar mecanismos para verificação de entrega. No entanto, o UDP contém um cabeçalho mais simples, com menos informações, fazendo com que seja mais “rápido” para a transmissão. Esse é um fator bem relevante em aplicações de tempo real, como o VoIP. (LIMA, 2016, p. 25).

O tema que será abordado a seguir, diz respeito à tecnologia Voip, que usa como base o UDP que é o protocolo da camada de transporte mais utilizado para tráfego de voz na atualidade.

2.2 VoIP (VOZ SOBRE IP)

VoIP (Voice over Internet Protocol), já teve muitos nomes, como Voice Over Broadband (VoBB), Internet Telephony, IP Telephony, e broadband Phone. De acordo com Voip do Brasil (2018), a tecnologia Voip foi desenvolvida em torno de 1995, e originalmente serviu como alternativa para chamadas de longa distância e internacionais. É uma tecnologia que permite a transmissão de voz sobre protocolo de Internet e é transmitida através de uma rede de computadores.

“A utilização crescente e maciça da Internet instigou o surgimento de uma série de novas tecnologias, muitas vezes, substituindo algumas já existentes, como no caso do VoIP. A sigla VoIP tem origem em “Voz sobre IP”, ou seja, é uma tecnologia que permite que chamadas telefônicas sejam feitas por meio de uma conexão de banda larga, no lugar dos serviços de telefonia convencionais. O VoIP é um protocolo de redes, isto é, trata-se de normas e regras implementadas para que a voz saia de uma origem, seja dividida em pacotes, trafegue por redes de dados através do TCP/IP, chegue ao destino, os pacotes sejam reunidos e re-organizados, reconstruindo assim a voz para que esta seja reproduzida para o destino.” (KELLER, 2011, p.19).

Sob o ponto de vista de Ross (2007), “Dentre as muitas tecnologias convergentes, capazes de transportar voz e dados pela Internet, uma das que mais se destaca atualmente é a chamada Voz sobre IP ou simplesmente VoIP”. A voz é digitalizada e transformada em dados(bits), que são conduzidos pela rede de computadores em forma de pacotes:

“Nos serviços de telefonia IP, a voz passa por um processo de digitalização para que este possa viajar pela rede na forma de bits. Uma vez digitalizada, a voz é transmitida na forma de pacotes de dados usando o protocolo IP dentro de uma rede privativa ou rede onde há garantia do serviço oferecido.” (ROSS, 2007, p.8).

Segundo a Voip do Brasil (2018), “Em 2003, o Skype lançou seu software beta, e rapidamente ganhou atenção nacional. Por um lado, o Skype permitia que as pessoas fizessem chamadas de voz por computador de forma totalmente gratuita.”

O Voip proporciona diversos benefícios, porém são necessários alguns cuidados para que a tecnologia funcione corretamente e com um bom desempenho. A seguir serão apresentados alguns conceitos em relação ao principal protocolo

utilizado no meio VoIP, além fatores que influenciam na qualidade das chamadas, codificadores de áudio, e o software livre Asterisk.

2.2.1 SIP

O Protocolo de Iniciação de Sessão mais conhecido como SIP (Session Initiation Protocol) é um protocolo de código aberto, que utiliza o modelo “requisição-resposta”, similar ao HTTP, para iniciar sessões de comunicação interativa entre utilizadores. É um padrão da Internet Engineering Task Force (IETF), definida na RFC 2543 (1999). Pode-se entender que é através do protocolo SIP, que é realizada a comunicação e sinalização de início e finalização das chamadas Voip sendo que a voz propriamente dita, é transportada através de outro protocolo chamado RTP que iremos ver mais adiante.

Como afirma Keller (2011, p.27), para a implementação dos serviços com base em VoIP, é necessário existir algumas normas para a inicialização, estabelecimento e finalização da comunicação. Essas normas são chamadas de protocolos de comunicação. Os protocolos mais conhecidos são o SIP e o H.323. Fundamentalmente, ambos permitem aos usuários a mesma funcionalidade, ou seja, estabelecer uma comunicação multimídia.

Pode-se dizer que SIP trata-se de um protocolo de controle referente à camada de aplicações do Modelo de Referência OSI (Open System Interconnection), que é usado para iniciar, modificar ou terminar sessões ou chamadas multimídia entre usuários (RNP, 2018). As chamadas são iniciadas, modificadas e terminadas basicamente com os seguintes comandos:

- INVITE: Indica que o usuário está sendo convidado a participar de uma sessão multimídia.
- Trying: Indica que a requisição está sendo processada.
- 200 OK(Answer): Indica a confirmação de atendimento.

- ACK: Confirmação de mensagem recebida como resposta final a um INVITE.
- OPTIONS: Faz uma pergunta sobre quais métodos e extensões são suportados pelo servidor e pelo usuário descrito no campo de cabeçalho.
- BYE: Usado para liberar os recursos associados a uma ligação e forçar a desconexão/desligamento da mesma.
- CANCEL: Cancela uma requisição que ainda esteja pendente, ou seja, em andamento.

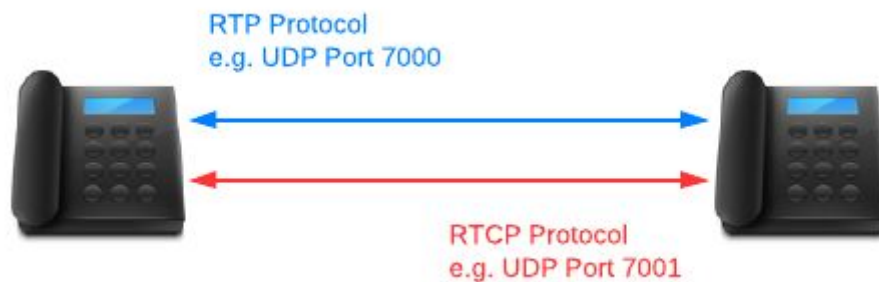
Através desses comandos também é definido em qual porta de comunicação será trafegada a voz de ambos os lados para que o protocolo RTP possa entrar em ação.

2.2.2 RTP

O protocolo de transporte de voz em tempo real, mais conhecido como RTP (Real-time transport Protocol), é conhecido por realizar o transporte de áudio em tempo real por meio de uma rede de computadores. Segundo Keller (2011), RTP é o responsável por fazer o transporte de voz, e a negociação dos formatos de mídia entre os participantes da comunicação. Deste modo, entendemos que o SIP depende do RTP para a transmissão dos pacotes de áudio que são trafegados na rede. A seguir iremos conhecer um pouco sobre o software Asterisk que tem o papel de uma central telefônica. Essa plataforma permite o uso de diversos protocolos de sinalização, inclusive o SIP, e o RTP.

O RTP abre duas portas para comunicação, uma para cada lado da ligação. Uma para o fluxo de mídia e outro para controle (feedback de QoS e controle de mídia). Os números das portas não são definidos com precisão, isso depende muito da aplicação e geralmente são randômicos. (VoipInfo, 2018)

Figura 2 : RTP



Fonte: (3CX, 2018)

O RTP é utilizado em conjunto com o RTP Control Protocol (RTCP). Enquanto ele conduz as transmissões de mídia (ex., áudio e vídeo), RTCP é utilizado para monitorar estatísticas e qualidade de serviço (QoS) e auxilia a transmissão de múltiplas transmissões. RTP é originado e recebido em números de portas iguais e comunicação RTCP associada utiliza o próximo maior número de porta. O RTP é uma das fundações do VoIP e é utilizado em conjunto com SIP que auxilia na configuração das conexões de rede. (3CX, 2018)

2.2.3 Codificadores

Para que seja possível a transmissão da voz e áudio pela rede de computadores, antes é necessário um processo de digitalização da voz. Esse processo é realizado através dos codificadores e decodificadores de áudio, chamados simplesmente "Codec".

"A digitalização da voz, ou a conversão do som analógico para sinais digitais é feita pelos Codecs (enCOde/DECode). Cada serviço, programa, telefone, gateway, equipamento VoIP suporta mais de um Codec e negocia qual será utilizado durante a inicialização das chamadas. Ao utilizar o VoIP, você deve escolher qual Codec será utilizado na comunicação. Essa escolha deve ter como base algumas premissas do seu ambiente, como: tamanho de banda disponível, capacidade de processamento necessária para realizar a digitalização da voz, quantidade de chamadas simultâneas a ser executada, entre outras não menos importantes." (KELLER, 2011, p.22).

Figura 2 - Características dos principais Codecs

| Codec | Banda (Kbps) | Payload (ms) | Banda nominal (ms) | Comentários |
|---------|--------------|--------------|--------------------|--|
| G.711 | 64 | 20 | 87.2 | Baixa utilização de CPU. |
| G.729a | 8 | 20 | 31.2 | Excelente utilização de banda e qualidade de voz. |
| G.722.1 | 48/56/64 | 30/30 | 160 | Excelente qualidade de áudio. É o considerado HD voicer CODEC. |
| G.723.1 | 5,3/6,3 | 30/30 | 20.8/21.9 | Exige muito poder de processamento. |
| G.726 | 24/32 | 20/20 | 47.2/55.2 | Baixo nível de compressão e de utilização de processamento. |
| GSM | 13 | | | Mesma codificação dos telefones celulares. |
| iLBC | 13.33/15 | 30/20 | | Resistente à perda de pacotes. |
| Speex | 8/16/32 | | | Utiliza taxa de bit variável para minimizar a utilização da banda. |

Fonte: Keller (2011)

É importante ressaltar, que os Codecs possuem características que influenciam diretamente na qualidade do áudio da chamada. A seguir serão apresentados outros aspectos que também influenciam na qualidade de VoIP.

2.2.4 Qualidade VoIP

Existem vários fatores que influenciam na qualidade de Voip, como o processamento de um computador que está executando um telefone virtual (Softphone) juntamente com outras aplicações pesadas por exemplo, um cabo de rede mal conectado ou ainda, a má qualidade ou instabilidade de uma rede sem fio wi-fi. Porém aqui serão elencados alguns fatores específicos e voltados à área de rede de computadores, que podem influenciar na qualidade de áudio de uma chamada feita através de VoIP, como a latência por exemplo sendo um indicador que permite medir o tempo de atraso da rede:

2.2.2.1 Latência (delay)

O bom funcionamento do VoIP, isto é, chamadas com áudio de qualidade, é totalmente dependente de alguns elementos considerados críticos. Como vimos

anteriormente, o áudio dentro da rede foi digitalizado, comprimido e dividido em pacotes para a sua transmissão de um ponto ao outro (KELLER, 2011). A latência por exemplo é um dos fatores que influenciam na qualidade das chamadas. A latência mede o tempo de atraso que um pacote leva para ir até o seu destino através da rede de computadores e voltar a origem. Se a latência for alta, acima de 50ms pode causar eco, e acima de 250ms sobreposição sinal:

“Dois problemas gerados pelo atraso dos pacotes numa comunicação fim-a-fim é o eco e a sobreposição de sinal (espécie de linha cruzada). Eco se torna um problema quando a demora de ida-e-volta for maior que 50 ms. Desde que o eco é percebido como um problema de qualidade, os sistemas de VoIP tem de prover mecanismos de cancelamento de eco. A sobreposição acontece se a demora for maior que 250 ms. O orçamento de demora de fim-para-fim é então o constrangimento principal e exigência motriz por reduzir demora por uma rede de pacote” (ROSS, 2007, p.14).

2.2.2.2 Jitter

Assim como a latência, o jitter também é um importante indicativo de qualidade de voz. Devido ao excesso de tráfego ou baixa largura de banda, o tempo de tráfego dos pacotes é diferente, e quanto maior a variação do tempo de tráfego dos pacotes, maior é o jitter. O excesso de jitter gera distorção no áudio da chamada, desde um pequeno chiado até o cancelamento da chamada em casos mais extremos(KELLER, 2011). Segundo Ross (2007), “O efeito jitter é a variação entre os tempos de chegada dos pacotes no destino provocados pela rede”.

De outro ponto de vista, jitter é o termo que define as variações de tempo da chegada do pacote de voz ao destino, ou seja, a variação do atraso fim a fim. Numa comunicação telefônica, os fluxos de pacotes de voz devem chegar ao destino numa harmonia constante, e de preferência no mesmo ritmo com que foram gerados pela origem. Se o jitter for muito grande, mesmo o atraso se mantendo dentro dos limites aceitáveis, a qualidade da comunicação vai decrescer até se tornar impossível. O jitter não pode ultrapassar 20 ms. (RIBEIRO, 2011)

2.2.2.3 Perda de pacotes

Na comunicação Voip não é aceitável a perda de pacotes, pois interfere diretamente na qualidade das chamadas, causando picotes e ligação muda causando transtorno para os usuários que dependem dessas ligações. Nas palavras de Ross (2007), “As redes IP não podem assegurar que todos os pacotes serão entregues, muito menos na ordem correta de envio. Alguns pacotes podem ser perdidos durante a transmissão quando a rede estiver congestionada”. “Perdas de pacote maior que 10% geralmente não são toleráveis” afirma Ross (2007).

O valor máximo de perdas de pacotes aceitável é de 1%. Ocorrendo uma perda de pacotes acima de 5% do total, uma conversa telefônica usando VoIP já fica comprometida (ROSS, 2007). Em suma, para que uma chamada não seja degradada, são diversos os fatores e características que influenciam diretamente na qualidade de voz para que esteja dentro dos padrões já conhecidos.

2.2.5 Asterisk

Asterisk é um software livre, de código aberto (OpenSource), que disponibiliza módulos e recursos encontrados em um PABX convencional, utilizando a tecnologia VoIP. Ele foi criado por Mark Spencer em 1999 (KELLER, 2011).

“O Asterisk PBX é, na minha opinião, uma revolução nas áreas de telefonia IP e PABX baseado em software. Durante muitos anos o mercado de telefonia foi ligado a equipamentos proprietários fabricados por grandes companhias multinacionais. [...] Com a entrada do Asterisk, mais e mais empresas vão poder experimentar recursos como URA – Unidade de resposta audível, DAC – distribuição automática de chamadas, mobilidade, correio de voz, e conferência, antes restritos à grandes companhias devido ao alto custo.” (GONÇALVES, 2005, p.3).

Na visão de Keller (2011, p. 19), “O Asterisk é considerado uma central telefônica híbrida, por implementar tanto as funções de uma central telefônica tradicional quanto os protocolos VoIP, ou seja, o Asterisk gerencia o áudio trafegado em canais de comunicação digitais, analógicos e também em redes TCP/IP”. Além de dispor de diversos benefícios e praticidade em sua configuração, também é

facilitado o uso de aplicações externas e integrações com outros sistemas por ser um software livre. No caso, será possível realizar facilmente a integração do Asterisk, com uma API online de reconhecimento de fala que o Google disponibiliza através de computação em nuvem.

2.3 RECONHECIMENTO DE FALA

Atualmente podemos citar os três mais conhecidos serviços de reconhecimento de voz: Google now, Microsoft Cortana, e Apple Siri. O reconhecimento de fala consiste em mapear um sinal acústico, capturado por um transdutor em um conjunto de palavras. (YNOGUTI, 1999, p. 8). Em suma, percebe-se que a tecnologia de reconhecimento de fala já vem sendo estudada há muito tempo, e gradativamente está sendo aperfeiçoada com novas técnicas:

“Uma ampla variedade de técnicas é usada para realizar o reconhecimento de fala. Existem muitos tipos de reconhecimento de fala. Existem muitos níveis de reconhecimento / análise / compreensão de fala. Normalmente, o reconhecimento de fala começa com a amostragem digital da fala. O próximo estágio é o processamento de sinais acústicos. A maioria das técnicas inclui análise espectral; por exemplo. Análise LPC (Codificação Preditiva Linear), MFCC (Coeficientes Cepstrais de Frequência Mel), modelagem da cóclea e muito mais.” (HUNT, 1997).

No entendimento de Ynoguti (1999), um sistema de reconhecimento de fala é capaz de converter e reconhecer um sinal acústico através de escolha de palavras a partir de um vocabulário finito.

“Com a tecnologia atual, a interatividade dos equipamentos eletrônicos vem crescendo bastante. Com cada vez mais poder computacional, smartphones, videogames, mais recentemente os televisores, e outros dispositivos, vem apresentando funcionalidades cada vez mais interessantes para o usuário, com o uso de filtros para fotos, realidade aumentada e reconhecimento de comandos por voz. (PAGANI et al, 2014).

2.3.1 Computação em Nuvem

Percebe-se que a computação em nuvem mudou e está cada vez melhorando e facilitando o mundo da tecnologia. Citando Velte (2017), “a computação em nuvem

é uma ideia que nos permite utilizar as mais variadas aplicações via internet, e em qualquer lugar e independente da plataforma, com a mesma facilidade de tê-las instaladas em nosso próprio computador.” Fisicamente, os dados armazenados em “Cloud” ou como chamamos nuvem, estão localizados em Centros de dados e processamento chamados “Data Centers”. Para Velte (2017), *data center* é um conjunto de servidores onde o aplicativo ou seus arquivos são armazenados. Poderia ser um grande quarto no porão de seu edifício ou em um quarto cheio de servidores no outro lado do mundo que você tenha acesso através da Internet.

“Podemos dizer que a computação em nuvem não é um produto de uma utilidade apenas. A infraestrutura pode ser implantada de várias maneiras diferentes. A infraestrutura dependerá do aplicativo e como o provedor escolheu construir a solução da nuvem. Esta é uma das vantagens-chave para utilizar a nuvem. Suas necessidades podem ser tão grandes que o número de servidores exigidos pode exceder sua necessidade ou orçamento para funcionar localmente. Por outro lado, você pode necessitar de pouca potência de processamento, assim você não precisa comprar um servidor dedicado para realizar o trabalho. A nuvem atende ambas as necessidades.” (VELTE, 2017).

É válido exemplificar aqui alguns aplicativos muito conhecidos para que se possa entender na prática o que é computação em nuvem:

- Dropbox e Google Drive (Armazenamento de arquivos: imagens, músicas, documentos, etc)
- Netflix, Spotify e Deezer (Filmes e Música online)
- Facebook (Rede social)
- Uber (Aplicativo de transporte)
- Whatsapp (Aplicativo de mensagem instantânea)

2.3.2 Google Cloud Speech Recog API

Como visto anteriormente, existem diversos aplicativos e serviços em nuvem disponíveis no mercado. O sistema de reconhecimento de fala do Google (Cloud Speech Recog API) também é um ‘aplicativo’ na nuvem ao qual é possível integrar com o PBX IP Asterisk através da Internet. Conforme afirma a empresa Google

(2018), “Cloud Speech API” é um sistema de conversão de voz em texto com tecnologia de aprendizado de máquina”.

Em concordância com Meeker (2017), o reconhecimento de fala do Google já é tão bom quanto o ouvido humano. O reconhecimento de voz está cada vez mais efetivo, e os resultados revelaram que a ferramenta obteve taxa de precisão de 95%, o que a deixa próxima do ouvido humano. De acordo com Meeker (2017), “O Google agora é capaz de entender a linguagem humana com precisão de 95% graças aos algoritmos de machine learning, que podem detectar a fala e responder com resultados significativos”.

“Reconhecimento de fala avançado. Com a Google Cloud Speech API, os desenvolvedores podem converter áudio em texto aplicando modelos de redes neurais avançados em uma API fácil de usar. A API reconhece mais de 110 idiomas e variantes para oferecer suporte à sua base de usuários global. É possível transcrever a voz de usuários por meio do microfone em um aplicativo, ativar o controle e comando de voz ou transcrever arquivos de áudio, entre muitos outros casos de uso. Reconhece o áudio enviado na solicitação e faz o armazenamento no Google Cloud Storage usando a mesma tecnologia que o Google usa nos próprios produtos.” (GOOGLE, 2018).

O Google Cloud Speech API funciona através de computação em nuvem.

3 PROCEDIMENTOS METODOLÓGICOS

O tema abordado no presente capítulo diz respeito aos procedimentos metodológicos utilizados para a estruturação deste trabalho. Serão descritos os métodos científicos utilizados para sua elaboração, e observados de acordo com a sua natureza, fins, meios, e procedimentos empregados do processo.

Método científico é o conjunto de processos ou operações mentais que devem ser empregados na investigação. É a linha de raciocínio adotada no processo de pesquisa.

O método aplicado ao presente trabalho será o método dedutivo, com uma abordagem quantitativa. Este método de acordo com o entendimento clássico, é o método que parte do geral e, a seguir, desce ao particular (Prodanov, Freitas, 2013).

“Por intermédio de uma cadeia de raciocínio em ordem descendente, de análise do geral para o particular, chega a uma conclusão. Usa o silogismo, a construção lógica para, a partir de duas premissas, retirar uma terceira logicamente decorrente das duas primeiras, denominada de conclusão.” (Prodanov, Freitas, 2018).

Semelhantemente, na concepção de (Gil 2008, p.9) “O método dedutivo, é o método que parte do geral para o particular. A partir de princípios, leis ou teorias consideradas verdadeiras e indiscutíveis, estudando ocorrências em casos particulares com base na lógica.”

Quanto à abordagem do problema, esta é definida como sendo quantitativa. No conceito de Kauark, Manhães e Medeiros (2010), pode ser enquadrado neste tipo de abordagem aquilo que pode ser quantificável, o que pode se traduzido em números opiniões e informações para classificá-las e analisá-las. Tal abordagem, requer o uso de recursos e de técnicas numéricas.

Segundo Fonseca (2002), a pesquisa científica é o resultado de um inquérito ou exame minucioso, realizado com o objetivo de resolver um problema, recorrendo a procedimentos científicos. Isto posto, cabe observar que do ponto de vista de sua natureza, o tipo de pesquisa aqui utilizada é de natureza aplicada. Assim, pode-se observar que este tipo de pesquisa procura gerar conhecimento a partir de uma visão prioritariamente no campo prático, voltados aos interesses de realidades locais (KAUARK, MANHÃES E MEDEIROS, 2010).

Na percepção de Gil (2008), a pesquisa exploratória, tem por principais características a exploração de assuntos pouco abordados, assim, tendendo a apresentar uma visão geral do problema e constituem geralmente a primeira etapa de uma investigação.

O objeto de pesquisa analisado sob a metodologia descritiva ocorre de maneira a descrever as características dos fenômenos estudados, utilizando-se de instrumento de coleta como o questionário e a observação sistemática, ou seja, sem interferência de opinião pessoal, estabelecendo assim as relações entre as variáveis. (KAUARK, MANHÃES E MEDEIROS, 2010).

Conforme relatado por Gil (2008), a pesquisa descritiva possui sua principal característica situada no uso de técnicas padronizadas de coleta de dados, tendo

assim por objetivo o estudo das características de determinado grupo, a partir do levantamento de informações obtidas pelos instrumentos de coleta.

Em relação ao procedimento utilizou-se o método experimental e bibliográfico. O método experimental consiste, especialmente, em submeter os objetos de estudo à influência de certas variáveis, em condições controladas e conhecidas pelo investigador, para observar os resultados que a variável produz no objeto (GIL, 2008).

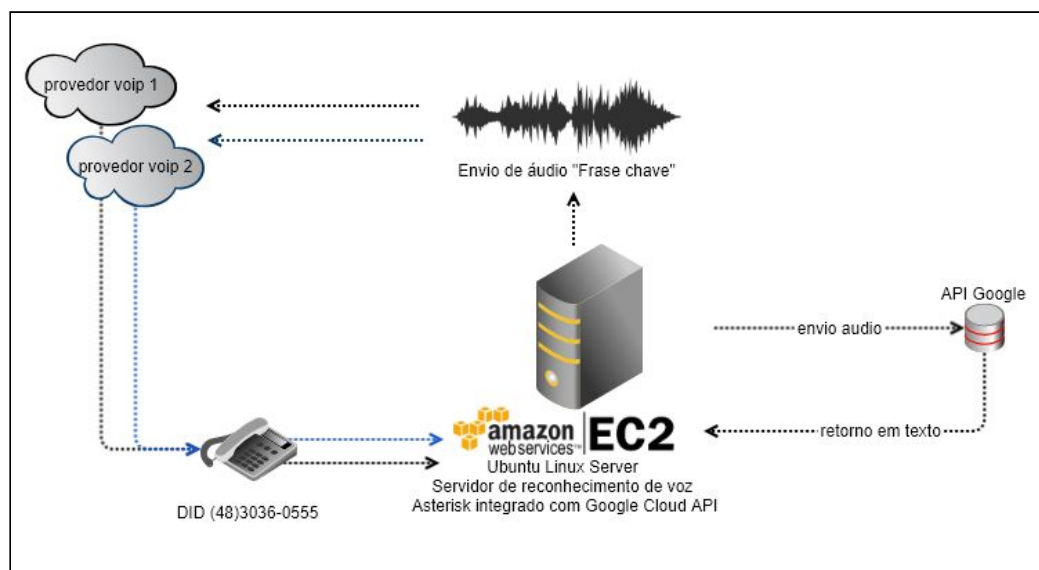
Para Prodanov e Freitas (2013, p.55) “A pesquisa bibliográfica é feita a partir do levantamento de referências teóricas já analisadas, e publicadas por meios escritos e eletrônicos, como livros, artigos científicos, páginas de web sites. De maneira semelhante, Fonseca (2002, p.32) afirma que “Qualquer trabalho científico inicia-se com uma pesquisa bibliográfica, que permite ao pesquisador conhecer o que já se estudou sobre o assunto”.

A seguir será apresentado o cenário proposto para a investigação do problema de pesquisa.

3.1 CENÁRIO PROPOSTO

Para desenvolver o cenário de testes e realizar a análise das informações obtidas, serão necessários alguns softwares e ferramentas. O servidor que fará o envio, a análise e o reconhecimento de voz será uma instância (máquina virtual) do Linux Ubuntu Server instalada e hospedada na Amazon AWS EC2. Nesse ambiente, será instalado o sistema Asterisk o qual fará o registro de um número virtual chamado de DID (Direct Inward Dialing) para que seja possível o recebimento de ligações.

Figura 3 - Cenário proposto



Fonte: Do autor (2018)

3.2 SOFTWARES UTILIZADOS

3.2.1 Amazon AWS

A Amazon Web Services (AWS) é uma plataforma de serviços em nuvem segura, oferecendo poder computacional, armazenamento de banco de dados, distribuição de conteúdo e outras funcionalidades para ajudar as empresas em seu dimensionamento e crescimento (AMAZON, 2018). Já o EC2 (Amazon Elastic Compute Cloud) é uma parte central da plataforma de cloud computing da Amazon AWS. O EC2 permite que os usuários aluguem computadores virtuais nos quais rodam suas próprias aplicações. O EC2 permite a implantação de aplicações escaláveis ao prover um Web service através do qual um usuário pode iniciar uma "Amazon Machine Image" para criar uma máquina virtual, que a Amazon chama de "instância", contendo qualquer software desejado. Um usuário pode criar, lançar e terminar instâncias do servidor, conforme necessário, pagando por hora pelos servidores ativos, daí o termo "elástico" (LAMONICA, 2008). Através do serviço EC2 da AWS, será instalado a instância do "Linux Ubuntu Server 16.04 LTS" para que seja possível montar a estrutura dos testes.

3.2.2 Instância Linux Ubuntu Server

Ubuntu é um sistema operacional de código aberto, construído a partir do núcleo Linux, baseado no Debian e utiliza GNOME como ambiente de desktop e servidor de sua mais recente versão com suporte de longo prazo (CANONICAL, 2018). Ele será utilizado como base para instalação de outros sistemas como Asterisk, e o desenvolvimento das ferramentas de análise.

3.2.3 Asterisk

Com esse software será possível realizar o registro de um número telefônico virtual(DID) para recebimento de chamadas através da rede de telefonia. Além disso será realizada a instalação e integração de uma API de reconhecimento de fala no Asterisk que quando retornar os resultados do reconhecimento chamará um Shell script que fará o tratamento dos dados e a pontuação que será abordada a seguir. Na prática, é o Asterisk que usa o protocolo SIP chamado "chan_sip" para realizar o registro e envio/recebimento das chamadas.

3.2.4 DID

O número virtual, mais conhecido como DID (Direct Inward Dialing) é uma configuração oferecida por companhias telefônicas para serem usadas no recebimento de ligações no sistema PBX de seus clientes, através do qual a companhia telefônica disponibiliza um número ou um conjunto de números associados com uma ou mais linhas telefônicas conectadas a PSTN (3CX, 2018). Neste trabalho será utilizado um provedor de DID que disponibiliza o número através de VoIP, para que seja possível receber ligações no servidor da Amazon.

3.2.5 Shell script

Com o intuito de registrar, automatizar os dados obtidos pela API do Google, e realizar uma pontuação de forma automática, será utilizado a linguagem de programação Shell Script por ser mais simples e prática de integrar com o plano de discagem do Asterisk. Segundo JARGAS(2004), Shell é o “prompt” da linha de comando do Linux que recebe os comandos digitados pelo usuário e os executa. Shell script é considerada uma linguagem de programação de script usada no sistema operacional Linux e MAC OS, com diferentes dialetos, dependendo do interpretador de comandos utilizado.

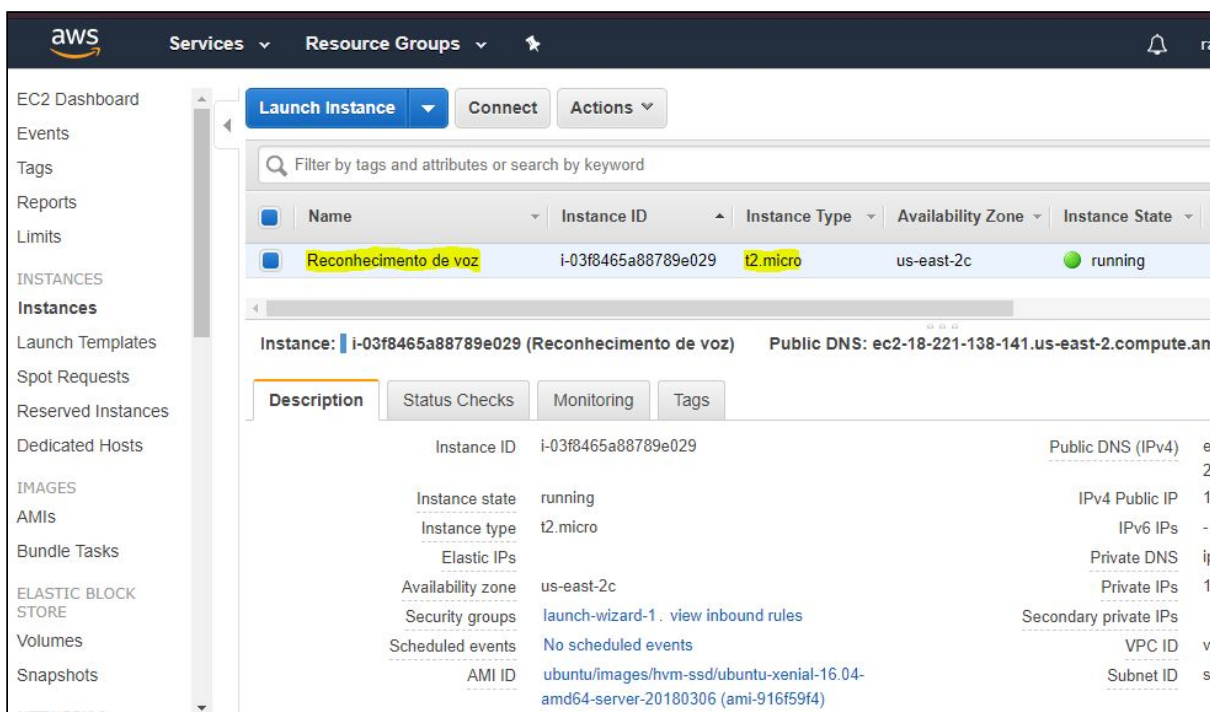
“O Shell é muito mais poderoso [...] Além dos comandos básicos para navegar entre diretórios e manipular arquivos, ele também possui todas as estruturas de uma linguagem de programação, como IF, FOR, WHILE, variáveis e funções. Com isso, também é possível usar o Shell para fazer Scripts e automatizar tarefas.” (JARGAS, 2004).

4 CENÁRIO DESENVOLVIDO

4.1 AWS EC2

Inicialmente para montar o cenário de testes, foi criada uma instância do Ubuntu Server 16.04 LTS na Amazon AWS EC2 que permitiu a instalação do Asterisk, e o desenvolvimento de um plano de discagem integrado com a api do Google.

Figura 4 - AWS EC2



Fonte: Do autor (2018)

Através do painel web da aws, é possível criar instâncias de diversos tipos de sistemas operacionais em poucos segundos. O painel conta com recursos de firewall, e toda estrutura de monitoramento do serviço. Quando o tipo da instância é “t2.micro” não tem custos, ou seja pode utilizar sem precisar pagar nada por isso. Esse tipo de instância normalmente é utilizado para realizar testes, onde a capacidade de processamento é limitada com um núcleo de CPU Xeon E5, 1GB de memória RAM, e 8GB de SSD. Além disso, um ip público e dinâmico é gerado para que seja possível um acesso remoto ao servidor.

4.2 ASTERISK

Através do gerenciador de pacotes do sistema operacional, chamado de “APT” (Advanced Packaging Tool), foi realizada a instalação do Asterisk versão 13 e alguns outros softwares que são considerados “dependências” para o funcionamento da ferramenta “Automatic Speech Recognition”:

- asterisk

- python-pip - alternative Python package installer
- libflac-dev, flac
- python-numpy, python-scipy, python-dev, python-setuptools, libsndfile-dev

4.3 GOOGLE SPEECH RECOGNITION

A API de integração com o Google e instruções de instalação e configuração foi acessada através do link:

<https://github.com/ederwander/Asterisk-Google-Speech-Recognition.git>

Foram criadas algumas extensões dentro de contextos no arquivo de configuração do asterisk “/etc/asterisk/extensions.conf”. Extensões e contextos são termos utilizados no plano de discagem do asterisk para melhor organizar e estruturar as configurações.

4.4 CONTEXTOS E EXTENSÕES

O context “[from-trunk]” foi utilizado para receber as ligações entrantes da rede de telefonia, e executar a análise no áudio onde é executado o atendimento com o comando “Answer”, logo após a voz é analisada através do eagi “pahh.py” que retorna o reconhecimento da voz dentro da variável “GoogleUtterance”.

Por fim, é executado o Shell-Script “analise.sh” que faz o reconhecimento, comparação, contabilização das palavras detectadas pelo Google Speech to Text, e o registro do resultado em um arquivo separado por vírgulas. O script “analise.sh” será visualizado com mais detalhes a seguir.

Figura 5 - Análise.sh

```
[from-trunk]
; Entrada OID
exten => _554830360555,1,Answer()
exten => _554830360555,n,Noop(===== RECEBENDO CHAMADA DE: ${CALLERID(num)} - API GOOGLE =====)
exten => _554830360555,n,eagi(pahh.py)
exten => _554830360555,n,GotoIf(${EXISTS(${GoogleUtterance})}?hello:bye)
exten => _554830360555,n(hello),System(/usr/src/analise.sh "${CALLERID(num)}" "${GoogleUtterance}")
exten => _554830360555,n(bye),Hangup()
```

Fonte: Do autor (2018)

A figura a seguir demonstra o contexto “[toca-audio]”, que por sua vez, é responsável por reproduzir os arquivos de áudio: “audio1” e “audio2” quando a ligação for atendida pelo contexto destino.

Figura 6 - Contexto toca-áudio

```
[toca-audio]
exten => s,1,Noop(===== TOCANDO AUDIO: ${EXTEN} =====)
exten => s,n,Playback(/var/lib/asterisk/sounds/audio1)
exten => s,n,Wait(2)
exten => s,n,Playback(/var/lib/asterisk/sounds/audio2)
exten => s,n,Wait(3)
exten => s,n,Hangup()
exten => h,1,Hangup()
```

Fonte: Do autor (2018)

No contexto “[default]”, foram criadas três extensões chamadas de provedor1, provedor2, e interno. As extensões “provedor[1,2]”, foram utilizadas para a saída das ligações através dos fornecedores 1 e 2. A extensão “interno” foi utilizada para validação dos áudios internamente sem passar por nenhum provedor.

Figura 7 - Contexto default

```
[default]
exten => _provedor1,1,Noop(=== CHAMADA SAINTE PROVEDOR 1 ===)
exten => _provedor1,n,Set(CALLERID(num)=4833330111)
exten => _provedor1,n,Dial(SIP/PROVEDOR_1/554830360555)
same => n,hangup()

exten => _provedor2,1,Noop(=== CHAMADA SAINTE PROVEDOR 2 ===)
exten => _provedor2,n,Set(CALLERID(num)=4833330222)
exten => _provedor2,n,Dial(SIP/PROVEDOR_2/554830360555)
same => n,hangup()

exten => interno,1,Answer()
exten => interno,n,Noop(===== CHAMADA DE: ${CALLERID(num)} =====)
exten => interno,n,eagi(pahh.py)
exten => interno,n,GotoIf($[${EXISTS}(${GoogleUtterance})]?hello:bye)
exten => interno,n(hello),System(/usr/src/analise.sh "interno" "${GoogleUtterance}")
exten => interno,n(bye),Hangup()
```

Fonte: Do autor (2018)

4.5 SHELL-SCRIPT “GENCALL_X.SH”

As chamadas foram geradas através de três Shell-scripts. Um para cada provedor e um para as chamadas Internas.

- A. gencall_interno.sh
- B. gencall_provedor1.sh
- C. gencall_provedor2.sh

Figura 8 - Gencall interno.sh

```
#!/bin/bash

calltmp="/tmp/call_${RANDOM}_tmp.call"

echo "Channel: Local/interno" > $calltmp
echo "CallerID: testel" >> $calltmp
echo "MaxRetries: 2" >> $calltmp
echo "RetryTime: 40" >> $calltmp
echo "WaitTime: 40" >> $calltmp
echo "Context: tocaaudio" >> $calltmp
echo "Extension: s" >> $calltmp
echo "Priority: 1" >> $calltmp

mv $calltmp /var/spool/asterisk/outgoing/
```

Fonte: Do autor (2018)

Os arquivos de script possuem a mesma estrutura, apenas as variáveis “Channel” e “CallerID” que são alteradas para que a chamada seja direcionada para o provedor corretamente através dos contextos e extensões que foram criadas anteriormente.

Figura 9 - gencall provedor1.sh

```
#!/bin/bash

calltmp="/tmp/call_${RANDOM}_tmp.call"

echo "Channel: Local/provedor1" > $calltmp
echo "CallerID: provedor1" >> $calltmp
echo "MaxRetries: 2" >> $calltmp
echo "RetryTime: 40" >> $calltmp
echo "WaitTime: 40" >> $calltmp
echo "Context: tocaaudio" >> $calltmp
echo "Extension: s" >> $calltmp
echo "Priority: 1" >> $calltmp
mv $calltmp /var/spool/asterisk/outgoing/
```

Fonte: Do autor (2018)

Os scripts foram configurados no agendador de tarefas do Linux para gerar as chamadas a cada um minuto. O agendador de tarefas é nativo do sistema operacional e é chamado de “cron” ou “crontab”.

Figura 10 - Agendador de tarefas do Linux (Cron)

```
# m h dom mon dow command
* 8 * * * /root/genccall_provedor2.sh
* 8 * * * /root/genccall_provedor1.sh
* 8 * * * /root/genccall_interno.sh
* 9 * * * /root/genccall_provedor2.sh
* 9 * * * /root/genccall_provedor1.sh
* 9 * * * /root/genccall_interno.sh
* 10 * * * /root/genccall_provedor2.sh
* 10 * * * /root/genccall_provedor1.sh
* 10 * * * /root/genccall_interno.sh
* 11 * * * /root/genccall_provedor2.sh
* 11 * * * /root/genccall_provedor1.sh
* 11 * * * /root/genccall_interno.sh
#
* 14 * * * /root/genccall_provedor2.sh
* 14 * * * /root/genccall_provedor1.sh
* 15 * * * /root/genccall_provedor2.sh
* 15 * * * /root/genccall_provedor1.sh
* 16 * * * /root/genccall_provedor2.sh
* 16 * * * /root/genccall_provedor1.sh
* 17 * * * /root/genccall_provedor2.sh
* 17 * * * /root/genccall_provedor1.sh
```

Fonte: Do autor (2018)

4.6 SHELL-SCRIPT “ANALISE.SH”

O script “analise.sh” é responsável por receber os dados(palavras) que o asterisk irá disponibilizar através da API do Google e fazer a análise, comparação e contabilizar o total de palavras corretas. Primeiramente foi definida uma variável que

recebe as palavras chaves do áudio. Através de um loop de repetição no código são analisadas cada palavra e pontuadas na variável “\$PONTUACAO” caso estejam de acordo. Por fim as informações são salvas em formato CSV (arquivo separado por pontos e vírgula) no diretório: “/tmp/resultados.txt”

Figura 11 - Parte do arquivo “/tmp/resultados.csv”

```
23/05/2018 11:18;25;004833330222 favor
23/05/2018 11:18;26;interno acho
23/05/2018 11:18;25;004833330111 por
23/05/2018 11:18;25;004833330222 acho
23/05/2018 11:19;26;interno acho
23/05/2018 11:19;25;+4833330222 acho
23/05/2018 11:19;27;004833330111 Boa
23/05/2018 11:20;26;+4833330222 acho
23/05/2018 11:20;25;+4833330222 acho
23/05/2018 11:20;26;004833330111 acho
```

Fonte: Do autor (2018)

O arquivo resultados.csv tem quatro colunas:

- Data no formato: Dia/Mês/Ano Hora:Minuto
- Pontuação (decimal)
- Número de origem para saber de qual provedor veio a ligação
- Resultados das Palavras detectadas pela API do Google

Figura 12 - Shell-script “analise.sh”


```

/bin/bash

PONTUACAO="0"

provedor=$1
/bin/echo ${@} > /tmp/resultadoGoogleUtterance.txt

# AUDIO 2
palavras_chaves="Boa tarde ketlin por favor a juliana do comercial é Pedro da totalvoice pra mim eu acho que ela
dela por meio dia estava dependendo retorno"

# AUDIO 1
$palavras_chaves="todos estavam estava falando rato foi até castelo roeu roupa vermelha rei roma enquanto ele do
carregado camelo camelos verde amarelo brasileiro brasileiros acordou"

for PalavraChave in $palavras_chaves; do

    cat /tmp/resultadoGoogleUtterance.txt|grep -iW ${PalavraChave}
    if [ $? == 0 ]; then
        PONTUACAO=$((PONTUACAO+1))
    fi
done

echo "`date +%d/%m/%Y %H:%M`"; ${PONTUACAO}; `cat /tmp/resultadoGoogleUtterance.txt`" >> /tmp/resultados.txt
~
~
~
~
~

```

Fonte: Do autor(2018)

4.7 Os ÁUDIOS

Os áudios foram convertidos e salvos no diretório: “/var/lib/asterisk/sounds/” no formato “WAV, PCM, 16 bit, mono, 8000 Hz” para que sejam compatíveis com o Asterisk.

A gravação do “audio1.wav” foi realizada com o microfone de um Notebook, onde o conteúdo haviam as seguintes palavras: “todos, estavam, estava, falando, rato, foi, até, castelo, roeu, roupa, vermelha, rei, roma, enquanto, dormia, sonhava, deserto, carruagem, carregada, carregado, camelo, camelos, verde, amarelo, brasileiro, brasileiros, e acordou”.

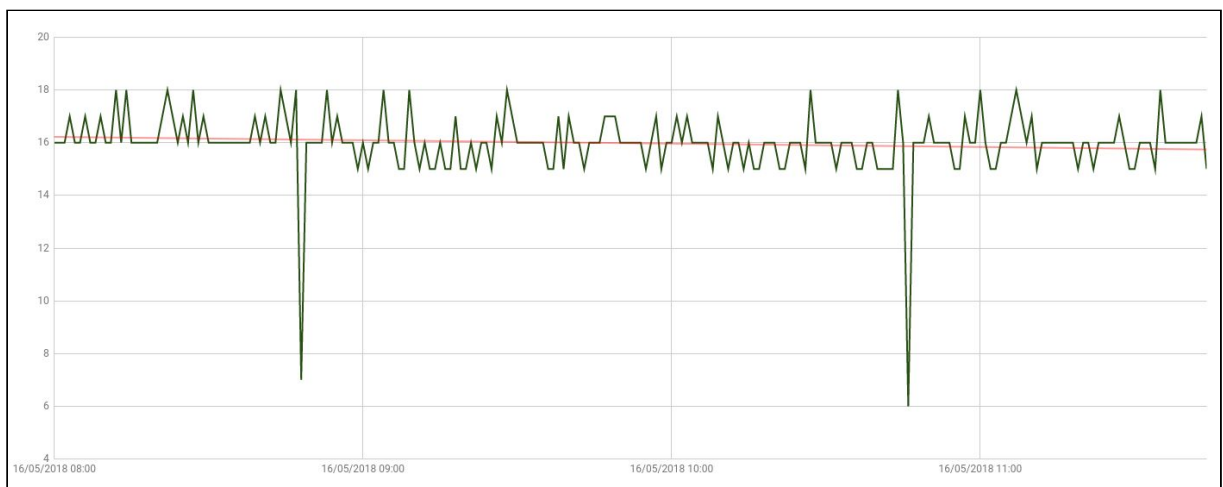
No entanto, o “audio2.mp3” foi uma gravação de conversação real entre duas pessoas onde existiam as seguintes palavras chaves: “Boa, tarde, ketlin, por, favor, juliana, comercia, Pedro, mim, acho, ela, autorizou, pedido, impressão, não, vou, precisar, mais, meio, dia, estava, dependendo, e retorno”

5 RESULTADOS

Os testes foram realizados em ambientes reais, e com provedores existentes no mercado. Após a coleta dos dados, foram analisados de forma independente os dados dos áudios 1 e 2 de cada provedor. Eles foram separados por períodos de “Manhã” e “Tarde”.

5.1 ÁUDIO 1

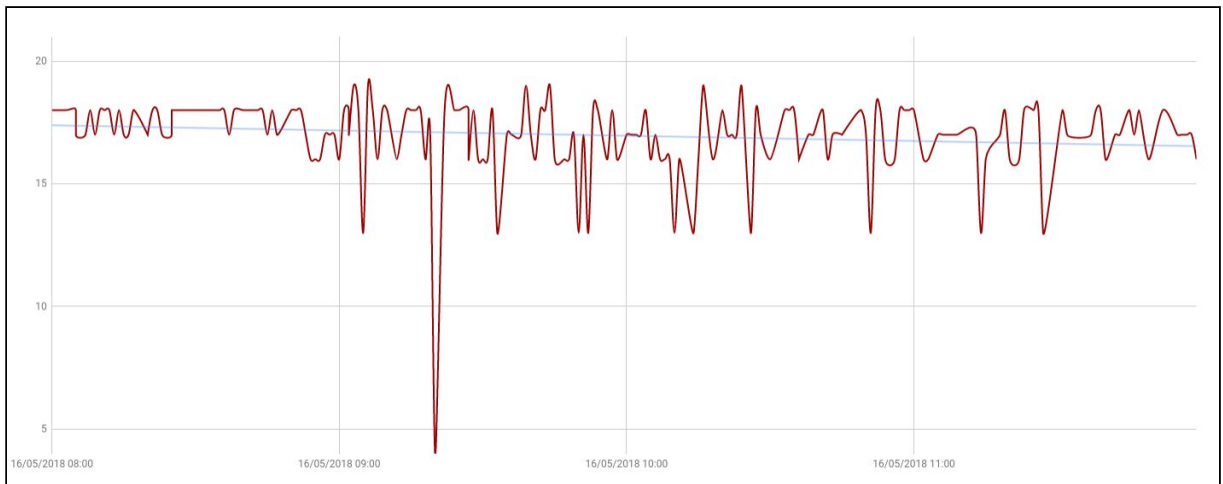
Figura 13 - Áudio 1, Provedor 1, Manhã



Fonte: Do autor (2018)

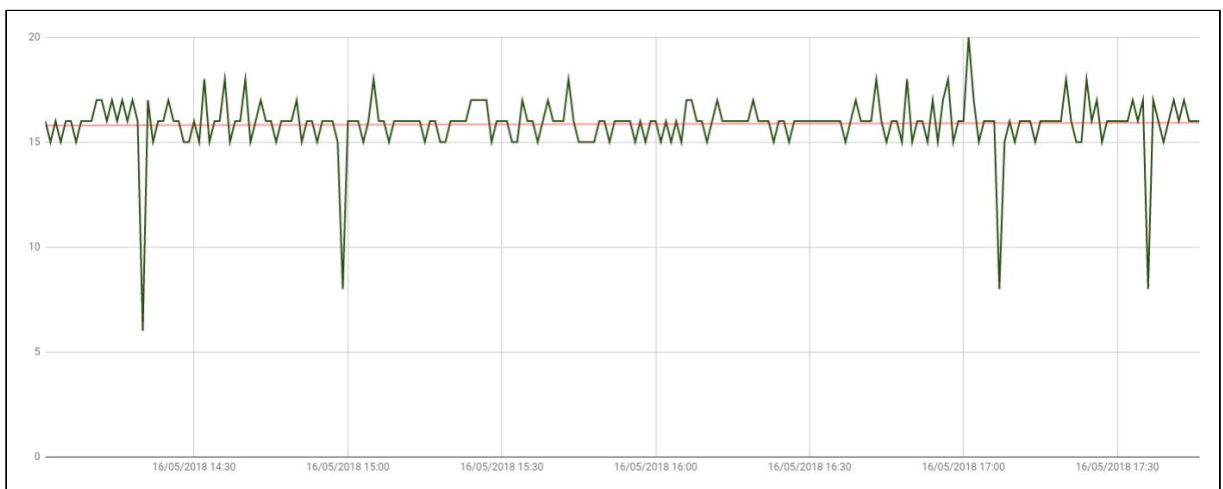
A figura 13 do provedor 1 apresentou uma média de detecção de 16 palavras com pouca variação durante o período da manhã, e uma pequena queda na detecção próximo às 8:48h e 10:46h. Enquanto a figura 14 do provedor 2, se mostrou mais instável com uma média de 17 palavras reconhecidas. Notou-se também uma queda na detecção de apenas 4 palavras as 9:20h.

Figura 14 - Áudio 1, Provedor 2, Manhã



Fonte: Do autor (2018)

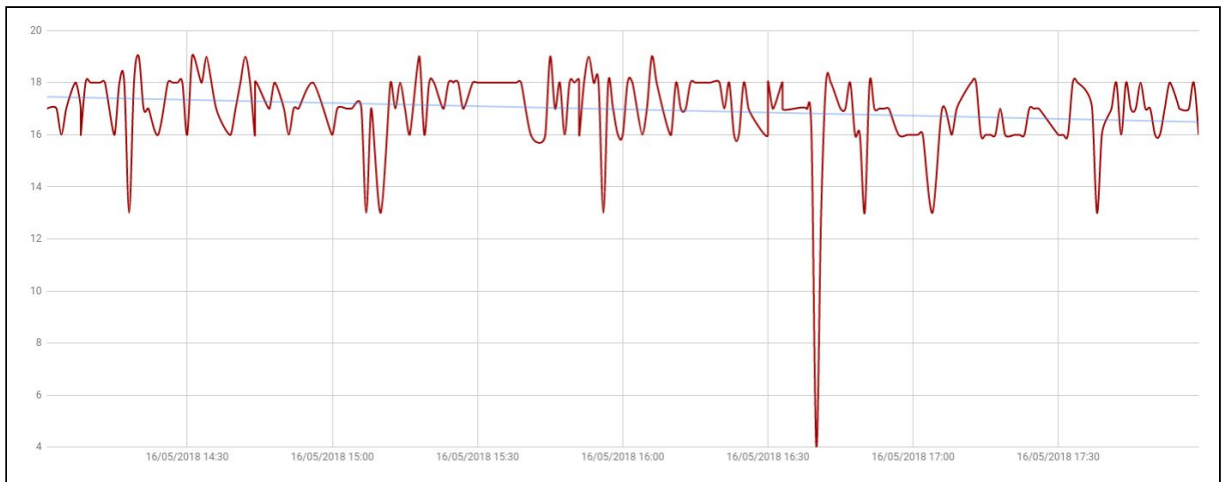
Figura 15 - Áudio 1, Provedor 1, Tarde



Fonte: Do autor (2018)

No período da tarde analisada na figura 15, nota-se que a média de detecção de palavras do provedor 1 permaneceu 16, porém com mais estabilidade quando comparado ao provedor 2 no mesmo período.

Figura 16 - Áudio 1, Provedor 2, Tarde

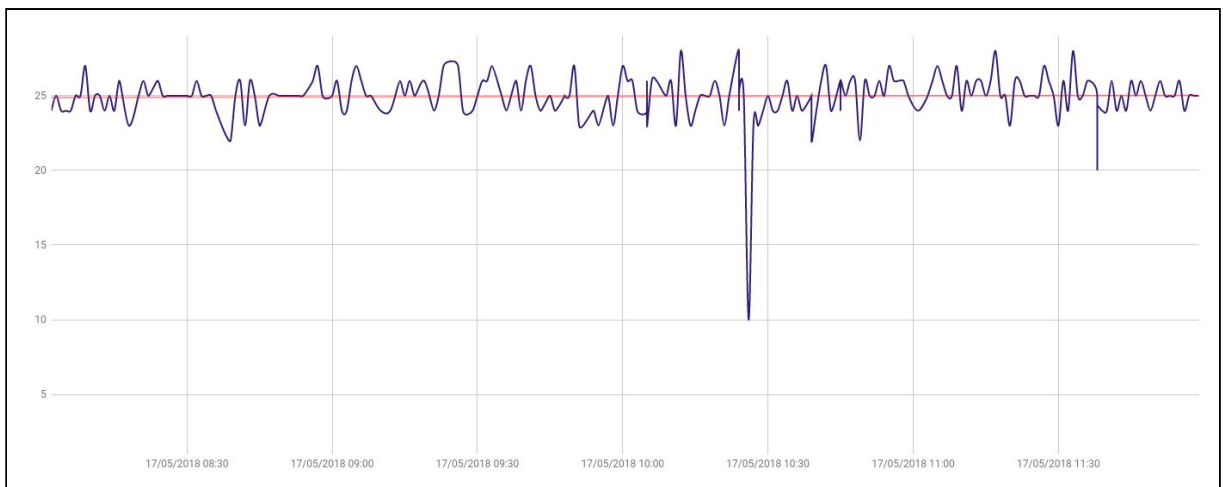


Fonte: Do autor (2018)

5.2 ÁUDIO 2

A figura 17 apresentou uma média de 25 palavras detectadas percebendo-se uma boa estabilidade na pontuação quando em comparada com a figura 18, do provedor 2 que apresentou uma pontuação média de 23 palavras.

Figura 17 - Áudio 2, provedor 1, manhã



Fonte: Do autor (2018)

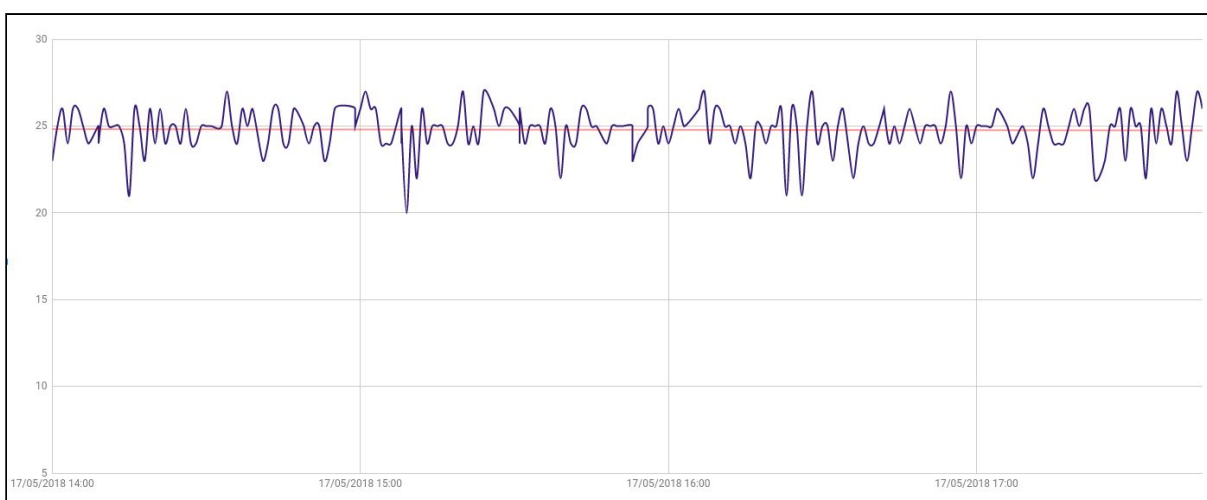
Figura 18 - Áudio 2, provedor 2, manhã



Fonte: Do autor (2018)

No período da tarde, a baixa detecção de palavras no provedor 2 continuou com mais frequência quando comparado ao período da manhã como pode ser visto nas figuras 18 e 20. Já o provedor 1, no período da tarde se manteve com uma boa estabilidade sem nenhuma queda brusca no gráfico, sendo que os testes foram executados simultaneamente com as mesmas condições recursos de rede, Internet, largura de banda, servidor e CPU para ambos provedores.

Figura 19 - Áudio 2, provedor 1, Tarde



Fonte: Do autor (2018)

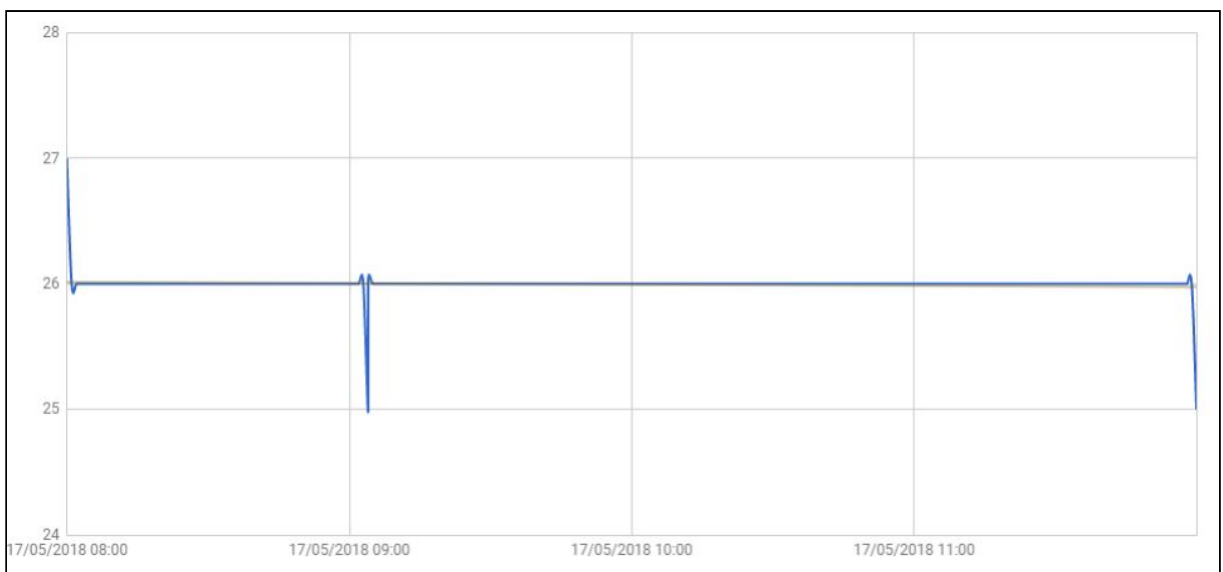
Figura 20 - Áudio 2, provedor 2, tarde



Fonte: Do autor (2018)

Tanto o áudio 1 quanto o áudio 2, quando enviado “diretamente” à API do Google, ou seja não passando por nenhuma rede de terceiros, a pontuação de detecção das palavras se mantiveram sempre muito estável apenas com baixas variações entre 25 e 27 palavras. Esse teste foi realizado com o objetivo de se ter uma base do que seria a pontuação ideal de detecção.

Figura 21 - Áudio 2, interno, manhã



6 CONSIDERAÇÕES FINAIS

Com a utilização e integração dos sistemas desenvolvidos no Asterisk e a API de reconhecimento de fala do Google nos cenários de testes da AWS, pode-se notar que o mecanismo trouxe uma linha de resultados estáveis em relação a pontuação principalmente do provedor 1, permitindo visualizar a queda na detecção de palavras e o acesso a um possível novo tipo de indicador para medir a qualidade em tempo real de um determinado provedor.

6.1 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Neste trabalho foram descritos assuntos que envolvem a qualidade e o monitoramento de chamadas VoIP em tempo real através de um sistema de reconhecimento de palavras. A recomendação é que seja desenvolvida uma plataforma onde os dados sejam exibidos de forma mais amigável ao usuário e que permita o envio de alertas via SMS, e-mail, etc, permita que o administrador de sistemas possa ser notificado que uma falha ocorreu e que está ocorrendo.

Um outro desafio talvez seja medir o outro lado da chamada, pois nesse cenário estamos “ouvindo” somente a origem da ligação. Além disso, pode-se realizar testes em outras redes

REFERÊNCIAS

AMAZON, Inc. **Amazon Web Services**: subtítulo. Disponível em: <<https://aws.amazon.com/pt/what-is-aws/>>. Acesso em: 04/06/2018.

BRITO, Samuel Henrique Bucke. **IPv6**: O Novo Protocolo da Internet. São Paulo: Novatec, 2013.

CANONICAL, Ltd. **The story of Ubuntu**. Disponível em: <<https://www.ubuntu.com/about>>. Acesso em: 04/06/2018.

COPPOLA, Giovanna. **Como será o Futuro da Indústria Voip**. Disponível em: <<http://blog.twsolutions.com.br/como-sera-o-futuro-da-industria-voip/>>. Acesso em: 04/06/2018.

GOOGLE, LLC. **CLOUD SPEECH API**: Conversão de voz em texto com tecnologia de aprendizado de máquina. Disponível em: <<https://cloud.google.com/speech/>>. Acesso em: 04/06/2018.

GONÇALVES, Flávio Eduardo de Andrade. **Asterisk PBX**: Como Construir e Configurar um PABX com Software Livre. Florianópolis, VOffice, 2005. Disponível em: <<http://www.asterisk.xpg.com.br/asterisk/AsteriskLivro.pdf>>. Acesso em: 04/06/2018.

HUNT, Andrew. **Speech Recognition**: How is speech recognition performed. Disponível em: <<http://www.speech.cs.cmu.edu/comp.speech/Section6/Q6.2.html>>. Acesso em: 04/06/2018.

INTERNET ENGINEERING TASK FORCE. **SIP**: Session Initiation Protocol. Disponível em: <<https://www.ietf.org/rfc/rfc2543.txt>>. Acesso em: 04/06/2018.

INTERNET ENGINEERING TASK FORCE. **INTERNET PROTOCOL**: RFC 791. Disponível em: <<https://tools.ietf.org/html/rfc791>>. Acesso em: 04/06/2018.

ISO, International Organization for Standardization. **All About ISO**. Disponível em: <<https://www.iso.org/about-us.html>>. Acesso em: 04/06/2018.

JARGAS, Aurélio. **Introdução ao Shell Script**. Disponível em: <<https://aurelio.net/shell/apostila-introducao-shell.pdf>>. Acesso em: 04/06/2018.

KELLER, Alexandre. **Asterisk**: na prática. São Paulo: Novatec, 2011.

KOLB, Juliana Jenny. **Modelo OSI**: (Open Systems Interconnection). Disponível em: <<http://jkolb.com.br/modelo-osi-open-systems-interconnection/>>. Acesso em: 04/06/2018.

LAMONICA, Martin. **Amazon Web Services adds resiliency to EC2 compute service**. Disponível em:

<<https://www.cnet.com/news/amazon-web-services-adds-resiliency-to-ec2-compute-service/>>. Acesso em: 04/06/2018.

LIMA, Tatiane de Oliveira de. **Priorização de tráfego de voz através de Serviços Diferenciados**. 74 páginas. TCC (Redes de Computadores) – SENAI CTAI, Florianópolis, 2016.

MEEKER, Mary. **Internet Trends**: 2018. Disponível em: <<http://www.kpcb.com/internet-trends>>. Acesso em: 04/06/2018.

MORAES, Alexandre...[et al]. **Redes de Computadores**: Da Ethernet a Internet. São Paulo: Érica, 2003.

MORAIS, Carlos Tadeu Queiroz de...[et al]. **Conceitos Sobre Internet e Web**. Porto Alegre: UFRGS, 2012.

NEVES, Julio Cezar. **Shell Linux**: Programação. Rio de Janeiro: Brasport, 2010.

PAGANI, Diego Henrique...[et al]. **Um Sistema de Reconhecimento de Fala**: Para Ensino de Robótica Destinado ao Controle de Robôs Lego® Nxt 2.0. 9 folhas. (Robótica) – Universidade Tecnológica Federal do Paraná, Medianeira, 2014.

PRODANOV, Cleber; FREITAS, Ernani Cesar. **Metodologia do trabalho científico**: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico. Novo Hamburgo: Feevale, 2013.

RIBEIRO, Gláucia da Silva. **Voz sobre IP II**: A Convergência de Dados e Voz. Disponível em: <<http://www.teleco.com.br/tutoriais/tutorialvoipconv2/default.asp>> Acesso em: 04/06/2018.

ROSS, Julio. **VoIP**: Voz Sobre IP. Rio de Janeiro: Antenna, 2007.

RNP, Rede Nacional de Ensino e Pesquisa. **Introdução ao protocolo SIP**. Disponível em: https://eng.registro.br/inoc/SIP_iNOC.pdf Acesso em: 04/06/2018.

VELTE, Anthony T.; VELTE, Toby J. Ph.D.; ELSENPETER, Robert . **Cloud Computing**: Uma abordagem Prática. Rio de Janeiro: Alta Books, 2012.

VOIPdoBrasil. **A História do VoIP**. Disponível em: <https://www.voipdobrasil.com.br/blog/a_historia_do_voip>. Acesso em: 04/06/2018.

VoipInfo. **RTP**. Disponível em: <https://www.voip-info.org/rtp/>. Acesso em: 26/06/2018

YNOGUTI, Carlos Alberto. **Reconhecimento de fala contínua usando modelos ocultos**. 152 páginas. Tese Doutorado. UNICAMP, São Paulo, 1999.

3CX, **RTP**. Disponível em: <https://www.3cx.com.br/voip-sip/rtp/> Acessado em: 26/06/2018

APÊNDICES

ANEXOS