

Interceptações telefônicas — formatos digitais úteis para análise



Marcelo Stopanovski
Cientista de Dados Jurídicos

Na [coluna da quinzena passada](#) tratamos das interceptações telefônicas

judiciais, as quebras de sigilo telefônico, nos aspectos de seu funcionamento geral de dados e dos documentos que as conferem validade. A ideia do último artigo foi traçar um panorama do que pode ser solicitado ao decisor e que representa a íntegra do material a ser acessado pela defesa.

No texto que segue, trataremos do formato dos arquivos digitais esperados na resposta às possíveis ordens de entrega exaradas pelo juízo, tanto às operadoras telefônicas, como aos investigadores.

Partindo-se do pressuposto que todas as informações estão agora disponíveis no local responsável por gerir as informações do processo, pois um pedido completo foi feito ao juízo e este o deferiu. Este local pode ser uma vara judicial ou um tribunal, destacando-se que os autos digitais do processo eletrônico compõem somente uma parte do material esperado, especialmente não contém os metadados e os áudios.

A defesa levará um HD (*Hard Disk Drive*) para este local e pedirá cópia integral do material. Um HD padrão de mercado com 750GB (*Gigabytes*) estará quase sempre de bom tamanho. O ideal é que o local fonte forneça um *Hash* certificador da autenticidade e integridade da totalidade dos dados fornecidos. Ressaltando que os áudios e os metadados, por exemplo, deverão vir com seus próprios certificados desde sua origem. Conversamos um pouco sobre essa tecnologia de certificação em um [artigo anterior sobre e-mails](#).

O formato digital das fontes de dados deve ser adequado às funções a que se destinam. Cada tipo de dado possui um formato melhor adaptado à análise a ser efetuada sobre ele.

Os dados que possuem características de tabelas devem ser disponibilizados ou em texto separado por vírgulas (CSV) ou Microsoft Excel (XLS ou XLSx) ou Marcação genérica (eXtensible Markup Language – XML). Tais formatos são abertos, com exceção do MS Excel, e permitem que as mais variadas ferramentas os leiam e os importem, não fazendo diferença de como estejam guardados na



ferramenta original que os exportou. Todos esses formatos citados são tipos de texto simples, passíveis de pesquisa por busca de palavras.

Em [referência ao artigo anterior sobre interceptações](#), estes formatos são úteis no caso dos itens 1, 2 e 3 da Tabela 2, quais sejam: o extrato telefônico cobrado do cliente da linha interceptada, o histórico (log) de acesso ao cadastro de clientes por parte dos investigadores e as mensagens de texto (SMS) enviadas ou recebidas, todos estes itens disponíveis no âmbito da operadora de telefonia.

Quanto aos dados que se consubstanciam nos três tipos de relatório que a investigação deve disponibilizar para a defesa, tem-se:

A exportação para conferência (item 4 da Tabela 2 do [artigo anterior](#)) deve ser armazenada de forma que possa ter conferida sua autenticidade e integridade por meio de senhas e algoritmos (novamente o *Hash*, a exemplo de RDS, MD5 ou SHA3). O formato destes arquivos pode ser proprietário, ou seja, exclusivo da ferramenta de interceptação que exporta estes dados.

O importante é que a exportação para interoperabilidade (item 5 da Tabela 2 do [artigo anterior](#)) possa ser feita direto pela ferramenta de visualização das interceptações ou que guarde garantia que os dados são os mesmos. Neste tipo de exportação os formatos resultantes, que são geralmente os encontrados e servem para seu propósito, são: texto tabulado (TXT), formato de Internet simples (HTML) ou texto com formatação (RTF).

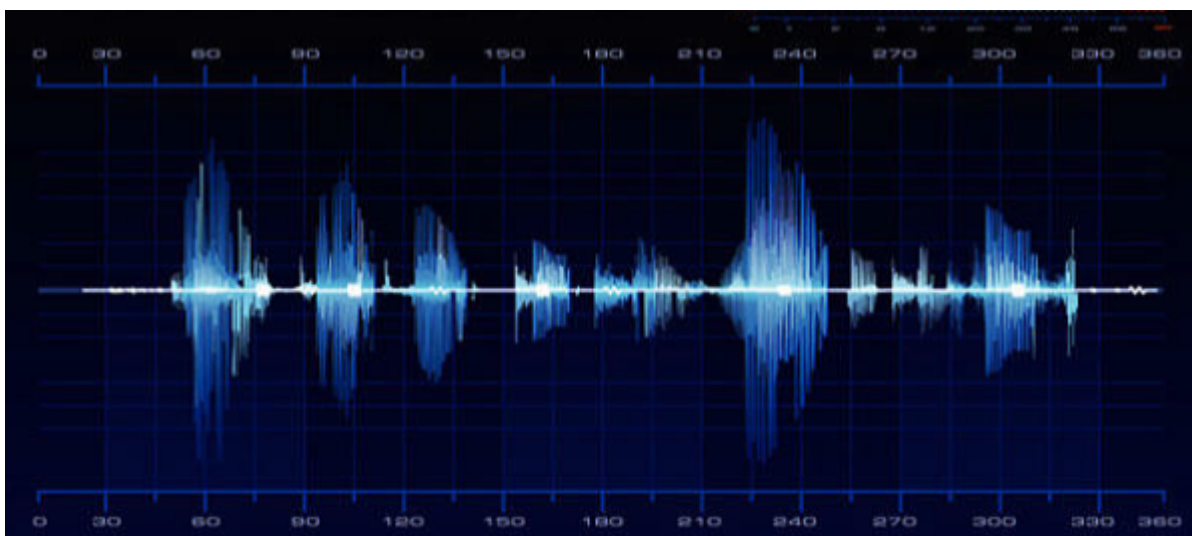
O ideal seria um formato XML com descrição padronizada entre as ferramentas de interceptação disponíveis no mercado, mas isso ainda não ocorre.

Nesta exportação para interoperabilidade além dos metadados (dados sobre as interceptações, como data e hora da captação), os áudios devem ser disponibilizados em formato WAV sem compressão.

Neste ponto cabe um comentário sobre os áudios e a possibilidade de que estes venham a ser submetidos a uma perícia. O campo do conhecimento que geralmente está ligado a uma perícia de arquivos de áudio é o da Engenharia Elétrica. O profissional perito irá analisar o sinal, a onda, de áudio e suas características, em busca de possíveis edições ou na tentativa de comparar o comportamento do sinal da fala da pessoa interceptada com a gravação de sua voz em outra fonte, visando validar se é ou não do mesmo falante.

A figura abaixo exemplifica um sinal de voz em um histograma (tipo de gráfico) comum nos programas de edição de som. Caso uma edição ou corte seja feito no áudio, será possível detectá-lo pelo comportamento do sinal.

Intel em Adrenaline.uol.com.br



Fonte: Intel em Adrenaline.uol.com.br

O formato necessário WAV sem compressão significa que todos os dados do áudio estarão disponíveis para a perícia. O uso de formatos comprimidos, os quais ocupam menos espaço para armazenamento ou transmissão, como o conhecido MP3, transformam o sinal, suprimindo, grosseiramente falando, repetições ou retirando trechos de silêncio. Como diria um colega: “Retiram o silêncio? Mas o silêncio em alguns casos diz muito!” Muitos músicos não escutam MP3 pois têm a sensação de que faltam notas.

Finalmente, quanto aos relatórios da investigação, o relatório de inteligência deve vir em formato *Microsoft Word* (DOC ou DOCx), e, caso contenha áudios embutidos, estes devem ser disponibilizados em pasta separada, também em formato WAV. Aqui o formato MS Word é útil para a localização e separação automática das ligações consideradas relevantes pela investigação, funcionalidades que seriam mais difíceis de serem utilizadas caso os arquivos estejam em um PDF.

O formato PDF (*Adobe Portable Document Format*) será o formato dos documentos dos autos, ressaltando que estes devem conter o texto passível de ser pesquisado. Deve ser possível localizar o nome do cliente dentro dos autos, por exemplo, utilizando a função de localizar texto do *Adobe Reader*.

Neste ponto cabe mais uma ressalva, é necessário o texto de forma pesquisável, passível de ser selecionado e copiado, para que seja possível processar os documentos de forma produtiva.

Uma digitalização dos autos, com o uso de um *scanner* ou câmera fotográfica, captura imagens do texto, mas não o texto em si de forma disponível para localização, seleção e cópia, para que isso aconteça a imagem deve ser submetida a uma tecnologia chamada OCR (*Optical Character Recognition*).

No processamento do OCR cada borrão na imagem é lido pelo computador e comparado a padrões de caracteres da língua em análise, resultado em um reconhecimento da letra. Essa tecnologia possui de forma inerente uma taxa de erros, visto que um texto escrito à mão ou uma assinatura que passou em cima do texto a ser reconhecido, não resultarão em reconhecimento do carácter a ser identificado.

Em resumo, o formato dos arquivos digitais determina as possibilidades de análise dos mesmos. O acesso à íntegra dos procedimentos de quebra de sigilo telefônico não é suficiente se o que for acessado



não puder ser processado dentro do estado da arte da tecnologia da informação.

Referências:

i-luminas. *Roteiro de coleta de dados e documentos em processos com quebra de sigilo telefônico*. Brasília, 2016. Disponível [aqui](#). Acessado em 31 de agosto de 2016.

Date Created

31/08/2016