SNIA

STORAGE NETWORKING INDUSTRY ASSOCIATION

EDUCATION

Business Intelligence From Mining VoIP Voice Recordings

Ray Lucchesi, Silverton Consulting, Inc.

SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individuals may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced without modification
  - The SNIA must be acknowledged as source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.

Abstract

Business Intelligence From Mining VoIP Voice Recordings

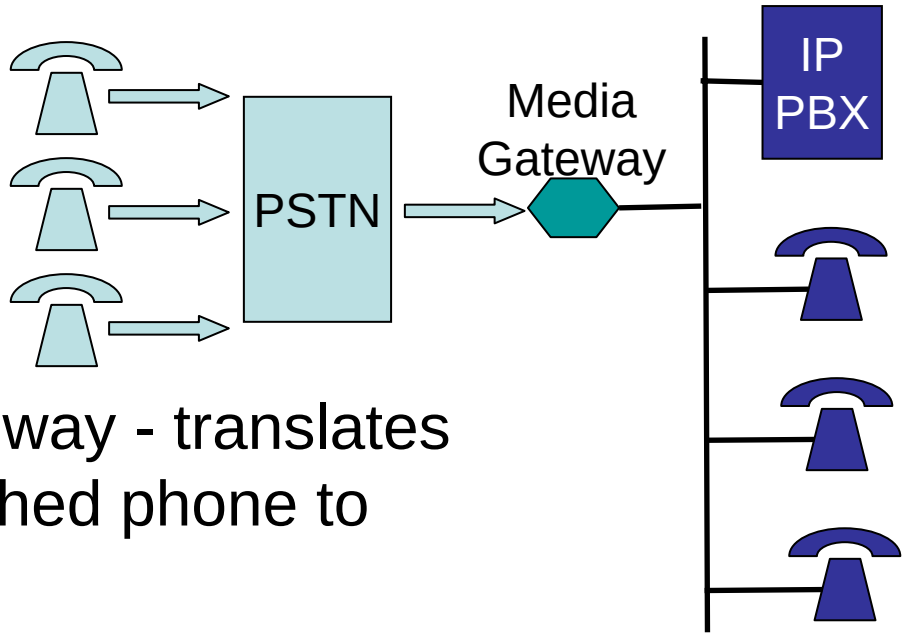
Today the convergence of telephony and the digital world due to the VoIP changeover is creating unprecedented opportunities to exploit and mine voice data not readily available before. Our presentation will describe technology used to record VoIP voice calls and detail typical storage architectures for voice storage. We conclude with how to mine voice recordings using keyword spotting, phonetic search, and other technology readily available today and discuss future voice recognition trends.

VoIP recording agenda

SNIA
STORAGE NETWORKING INDUSTRY ASSOCIATION
EDUCATION

- Voice recording architecture
- Voice recording storage
- Voice recording data mining

VoIP architecture



Media Gateway - translates from switched phone to VoIP

IP PBX - VoIP public branch exchange

VoIP adoption

- Predicting >50% of enterprise PBX's to be VoIP by 2008
- Y2K caused companies to upgrade call centers - Many now coming off lease/fully depreciated
- VoIP latest technology
  - Inexpensive
  - Convergent to digital
  - QOS improving
  - Open standards available
  - Increasing vendor support

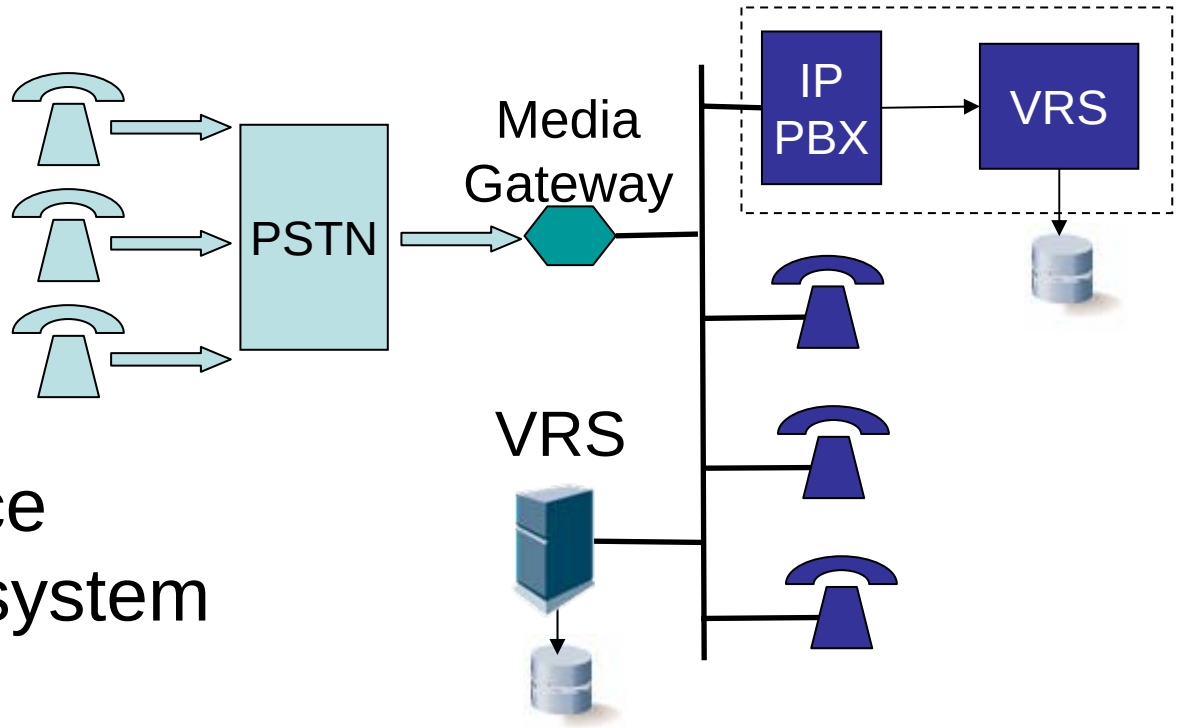
Why companies record voice

- Quality assurance
  - "…this call may be recorded for quality assurance purposes…"
- Compliance
  - DOT requires all transportation company complaints to be recorded and listened to within 6 months
- Data mining
  - Data mining requirements may change over time and can re-mine same data

Legal Concerns

- USA allows recording calls
  - 38 states if one party provides authorization, “one-party consent”
  - 12 states if all parties on the call provide authorization, “two-party consent”
    - California, Connecticut, Florida, Illinois, Maryland, Massachusetts, Michigan, Montana, Nevada, New Hampshire, Pennsylvania and Washington.
- International varies by jurisdiction

VoIP recording architecture



VRS - Voice recording system

VRS network attached

- VRS combines VoIP packets into call voice recording
- Configure voice traffic through one switch and VLAN
- Switch duplicates voice traffic to VRS server
  - VRS has packet sniffer at Span Port
  - Remote sites echo traffic back to switch
- Works with any IP PBX

VRS PBX attached

- Proprietary interface to VRS system or VRS onboard PBX
- Possibly proprietary voice recording format
- IP PBX proprietary data available

IP PBX functionality

- Auto attendant: dial by name, extension, or group, ACD, CCR
- Voicemail: security, multiple greetings & mailboxes, vmail review, forwarding, notification & email
- Conferencing: public & private, recording, conference admin.
- Others: call park, hold, & xfer, hunt groups, speed dial, remote ofc, call logs & detail records
- Music

VoIP protocols - ITU and IETF

- Gateway control: MGCP, SGCP,IPDC, MEGACO
- Signaling: H-323, H-225, H-235, H-245, H-450, T-38, T-120
- Session: SIP, SDP, SAP
- Media transport: RTP, RTCP, RTSP
- Media encoding: G.711, G.722, G.723, G.726, G.727, G.729

Contact center statistics

- Avg. contact center ~10M calls/year, 500-1000 operators
- Avg. call ~3-5 minutes, but length industry and call center driven
- Not unusual to have 5000 operator call center in Asia, US call centers smaller

VoIP network loading

- Depends on
  - CODEC, ranging from 5.3 to 64Kbps - impacts QOS
  - Number of active voice lines
- For 100 active lines
  - 1.6 Mbps using G.723.1 CODEC
  - 8.0 Mbps using G.711 CODEC

VoIP recording agenda

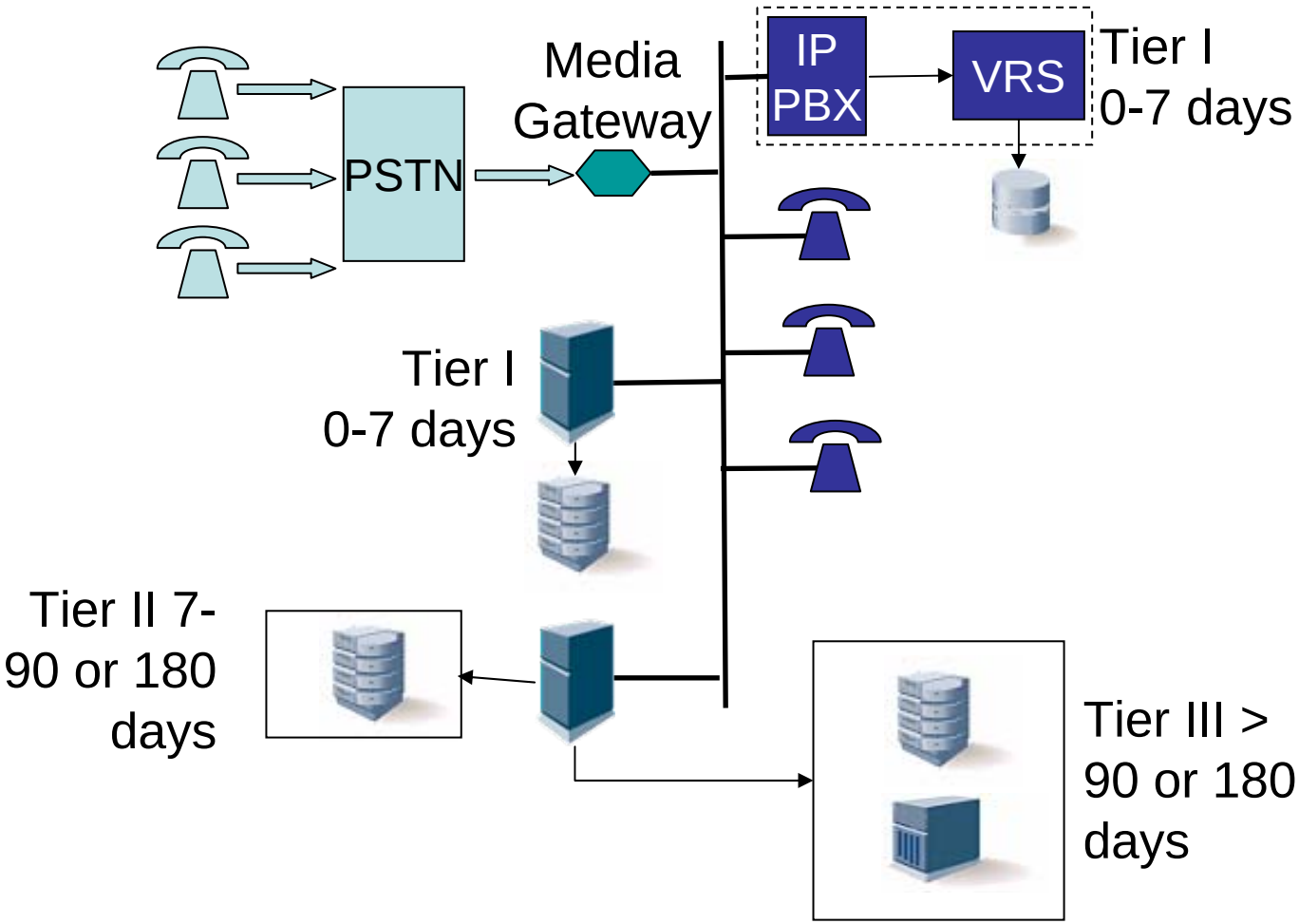
SNIA
Storage Networking Industry Association
EDUCATION

- Voice recording architecture
- Voice recording storage
- Voice recording data mining

Data center surprise

- Many start 100% call recording to enterprise class disk
- Quickly discover TB/month of disk tied up, then move to Tier II
- Still find ~10s TB/yr too expensive for Tier II then move to Tier III
- Finally move from saving 100% of calls to saving only select calls

VoIP storage tiers



VoIP voice recording

- File or database entries
- Raw file (1-15MB) depending on call length & VoIP CODEC
- VRS can convert CODEC to standard media format
  - .WAV, .MP3, etc.
  - ~1MB/call minute
- Created and read sequentially
- Tier I recordings done in real time

Call detail record

- Typically cut by IP PBX system
- Includes: caller-id, number called, time-date stamp, call duration, etc.
- Can include PBX info: operator-id, customer-id, operator login time, etc.
- Periodic screen shots of operator screen
- Not that large
- Used as cross index to voice recording

VoIP storage tier I

- 0-7 days
- Real time sequential write access
- For average call center
  - ~3.3TB of voice recordings/month
  - ~833K voice files or entries/month
  - ~833K call detail files/month

→ ∑ **~1.7M files/month**

VoIP storage tier II

- 7 to 90 or 180 days
- SAN or NAS storage
- Sequential, read mostly
- For QA ~1-3% of voice recordings accessed/month
- For Compliance up to 100% voice recordings accessed/month
- Discarding some un-needed files

VoIP storage tier III

- > 90 or 180 days
- Access rates un-quantifiable but << 1%/month
- Call metadata used to determine
  - How much to retain long term?
  - How long to retain voice recordings? - typical 7 yrs

VoIP voice SAN loading

- ~150GB/day recording load for Tier I
  - Real-time, high sequential write workload
- ~4.5GB/day QA-listening load for Tier II
- Add data movement - Tier I to II and Tier II to III
- Add compliance listening load for Tier II

Storage H/W

- Tier I - usually direct attached but can use enterprise class SAN storage
- Tier II - less expensive than enterprise class
- Tier III - archive appliances, tape library, or optical jukebox

VoIP disaster recovery considerations

- Analog PBX equipment caused distinct DR plans
- VoIP bringing call center DR plans back inside data center
- Higher network requirements to support VoIP

VoIP recording agenda

SNIA
STORAGE NETWORKING INDUSTRY ASSOCIATION
EDUCATION

- Voice recording architecture
- Voice recording storage
- Voice recording data mining

Voice recording data mining activities

- Identify call and business trends
  - Flood of calls traced to minor website change
- Improve business processes and products
  - Product trends detected sooner
- Tag calls for further review
  - Disgruntled customer scheduled for follow-up special offers
- Detect caller fraud
  - Biometric id of person calling based on prior calls
- Speech search engine

Speech Analytics

- Keyword spotting
  - Limited vocabulary identified
- Emotion analysis
  - Caller stress level assessed
- Talk pattern analysis
  - Call tempo analyzed

Speech keyword spotting

- Transcribes call into XML file using limited vocabulary
- Done mostly offline in post process
- Determines voice record processing options
- Experiments in real-time word spotting at state of the art call centers
- Used to spot product call trends or other keywords of interest

Voice emotion analysis

- Scores caller stress level
- Determines voice recording processing options
- Done in post processing step
- Experiments in real-time stress analysis at state of the art call centers
- Used to spot disgruntled/stressed callers

Talk pattern analysis

- Scores call talk pattern by identifying tempo, turn-taking, other conversation dynamics, and parts of call
- Determines voice recording processing options
- Done post processing
- Used mostly for quality assurance

Voice print identifier

- Biometric voice print from prior call general speech or from “pass phrase”
- Comparison made at next call to authenticate person speaking
- Done in real-time
- Used as a substitute PIN/Password or as alternate way to validate persons identity

Speech search engine

Based on vocabulary or phonetics

- Large vocabulary continuous speech recognition (LVCSR) advantages
  - Better accuracy
  - Transcription available
  - Storage efficient
- Phonetics advantages
  - Unlimited vocabulary
  - No re-processing for new keywords
  - Multi-language support
  - Faster indexing

Speech recognition

- Trained vs. untrained
- Continuous vs. paused speech
- Controlled vs. uncontrolled channel
- Interactive vs. passive
- Limited vs. unlimited vocabulary

→ **50-80% accurate today**

- Predicting better than human recognition by 2011
- Accuracy can improve through
  - Better software and systems
  - More processing power devoted to analyzing call

Q&A / Feedback

- Please send any questions or comments on this presentation to SNIA: trackapplications@snia.org

Many thanks to the following individuals for their contributions to this tutorial.

SNIA Education Committee

Raymond Lucchesi