Practical Attacks Against Encrypted VoIP Communications

HITBSECCONF2013: Malaysia

Shaun Colley & Dominic Chell

@domchell @mdseclabs

Agenda

- This is a talk about traffic analysis and pattern matching
- VoIP background
- NLP techniques
- Statistical modeling
- Case studies aka “the cool stuff”

Introduction

- VoIP is a popular replacement for traditional copper-wire telephone systems
- Bandwidth efficient and low cost
- Privacy has become an increasing concern
- Generally accepted that encryption should be used for end-to-end security
- But even if it's encrypted, is it secure?

Why?

- Widespread accusations of wiretapping
- Leaked documents allegedly claim NSA & GCHQ have some "capability" against encrypted VoIP
- *"The fact that GCHQ or a 2nd Party partner has a capability against a specific the encrypted used in a class or type of network communications technology. For example, VPNs, IPSec, TLS/SSL, HTTPS, SSH, encrypted chat, encrypted VoIP".*

MDSec

Previous Work

- Little work has been done by the security community
- Some interesting academic research
  - Uncovering Spoken Phrases in Encrypted Voice over IP Communications: *Wright, Ballard, Coull, Monrose, Masson*
  - Uncovering Spoken Phrases in Encrypted VoIP Conversations: *Doychev, Feld, Eckhardt, Neumann*
- Not widely publicised
- No proof of concepts

Background:
VoIP

VoIP Communications

- Similar to traditional digital telephony, VoIP involves signalling, session initialisation and setup as well as encoding of the voice signal
- Separated in to two channels that perform these actions:
  - Control channel
  - Data channel

Control Channel

- Operates at the application-layer
- Handles call setup, termination and other essential aspects of the call
- Uses a signalling protocol such as:
  - Session Initiation Protocol (SIP)
  - Extensible Messaging and Presence Protocol (XMPP)
  - H.323
  - Skype

Control Channel

- Handles sensitive call data such as source and destination endpoints, and can be used for modifying existing calls
- Typically protected with encryption, for example SIPS which adds TLS
- Often used to establish the the direct data connection for the voice traffic in the data channel

Data Channels

- The primary focus of our research
- Used to transmit encoded and compressed voice data
- Typically over UDP
- Voice data is transported using a transport protocol such as RTP

MDSec

Data Channels

- Commonplace for VoIP implementations to encrypt the data flow for confidentiality
- A common implementation is Secure Real-Time Transport Protocol (SRTP)
- By default will preserve the original RTP payload size
- "*None of the pre-defined encryption transforms uses any padding; for these, the RTP and SRTP payload sizes match exactly.*"

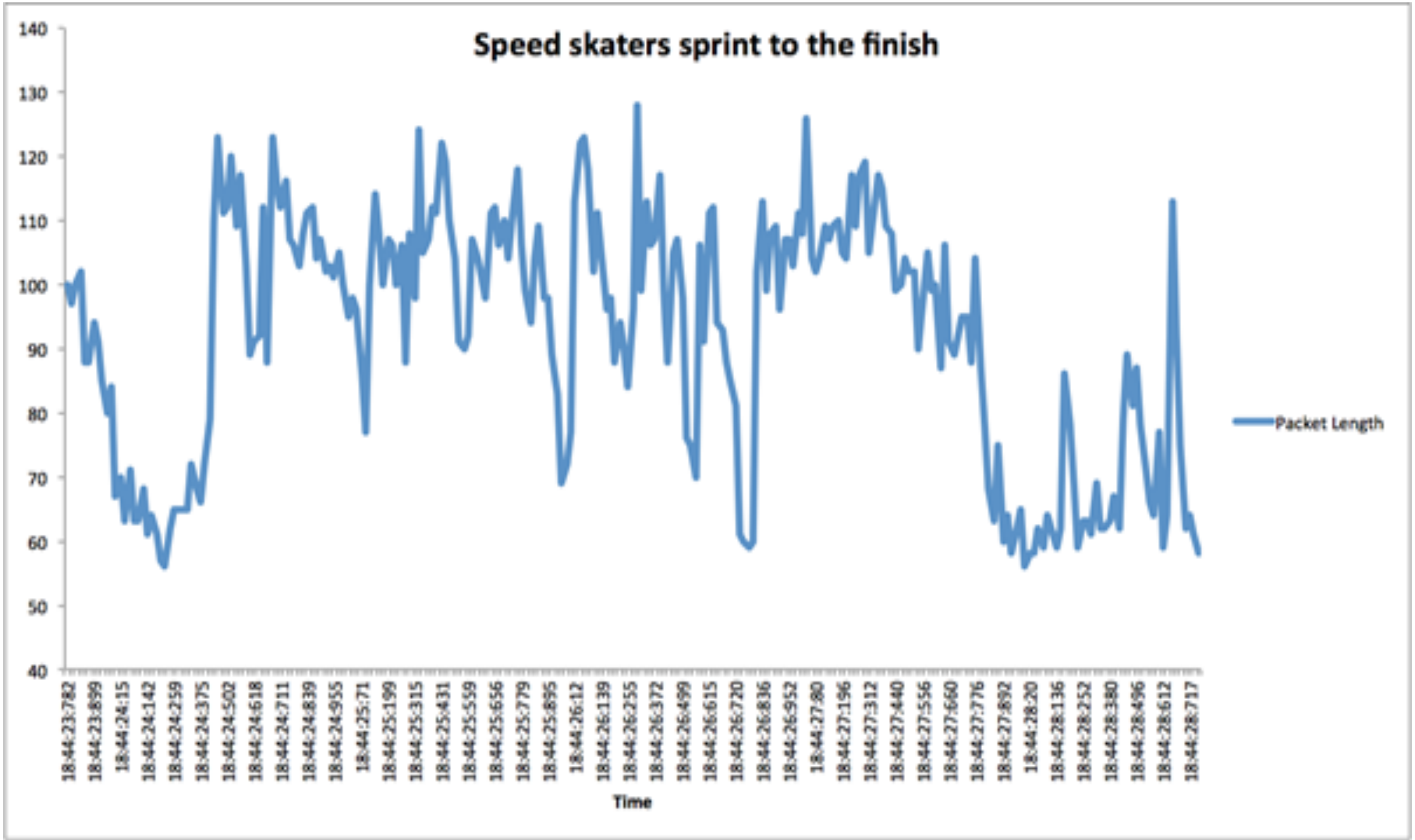
Background:
Codecs

Codecs

- Used to convert the analogue voice signal in to a digitally encoded and compressed representation
- Codecs strike a balance between bandwidth limitations and voice quality
- We're mostly interested in Variable Bit Rate (VBR) codecs

Variable Bitrate Codecs

- The codec can dynamically modify the bitrate of the transmitted stream
- Codecs like Speex will encode sounds at different bitrates
- For example, fricatives may be encoded at lower bitrates than vowels


Speed skaters sprint to the finish
140
130
120
110
100
90
80
70
60
50
40
18:44:23:782
18:44:23:899
18:44:24:15
18:44:24:142
18:44:24:259
18:44:24:375
18:44:24:502
18:44:24:618
18:44:24:711
18:44:24:839
18:44:24:955
18:44:25:71
18:44:25:199
18:44:25:315
18:44:25:431
18:44:25:559
18:44:25:656
18:44:25:779
18:44:25:895
18:44:26:12
18:44:26:139
18:44:26:255
18:44:26:372
18:44:26:499
18:44:26:615
18:44:26:720
18:44:26:836
18:44:26:952
18:44:27:80
18:44:27:196
18:44:27:312
18:44:27:440
18:44:27:556
18:44:27:660
18:44:27:776
18:44:27:892
18:44:28:20
18:44:28:136
18:44:28:252
18:44:28:380
18:44:28:496
18:44:28:612
18:44:28:717
Time
Packet Length

Variable Bitrate Codecs

- The primary benefit from VBR is a significantly better quality-to-bandwidth ratio compared to CBR
- Desirable in low bandwidth environments
  - Cellular
  - Slow WiFi

Background:
NLP and Statistical Analysis

Natural Language Processing

- Research techniques borrowed from NLP and bioinformatics
- Primarily the use of:
  - Profile Hidden Markov Models
  - Dynamic Time Warping

MDSec

Hidden Markov Models

- Statistical model that assigns probabilities to sequences of symbols
- Transitions from *Begin* state (B) to *End* state (E)
- Moves from state to state randomly but in line with transition distributions
- Transitions occur independently of any previous choices

Hidden Markov Models

- The model will continue to move between states and output symbols until the *End* state is reached
- The emitted symbols constitute the sequence

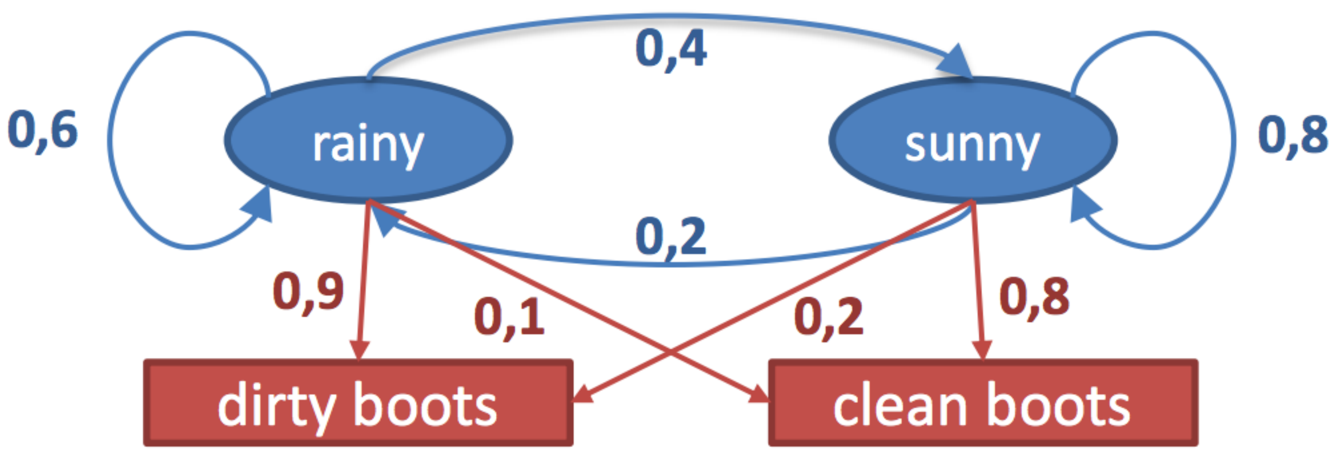


Image from http://isabel-drost.de/hadoop/slides/HMM.pdf

Hidden Markov Models

- A number of possible state paths from B to E
- *Best path* is the most likely path
- The Viterbi algorithm can be used to discover the most probable path
- Viterbi, *Forward* and *Backward* algorithms can all be used to determine probability that a model produced an output sequence

Hidden Markov Models

- The model can be "trained" by a collection of output sequences
- The Baum-Welch algorithm can be used to determine probability of a sequence based on previous sequences
- In the context of our research, packet lengths can be used as the sequences

Profile Hidden Markov Models

- A variation of HMM
- Introduces *Insert* and *Deletes*
- Allows the model to identify sequences with *Inserts* or *Deletes*
- Particularly relevant to analysis of audio codecs where identical utterances of the same phrase by the same speaker are unlikely to have identical patterns

Profile Hidden Markov Models

- Consider a model trained to recognise:

  A B C D

- The model can still recognise patterns with ***insertion***:

  A B X C D

- Or patterns with ***deletion***:

  A B C

Dynamic Time Warping

- Largely replaced by HMMs
- Measures similarity in sequences that vary in time or speed
- Commonly used in speech recognition
- Useful in our research because of the temporal element
- A packet capture is essentially a time series

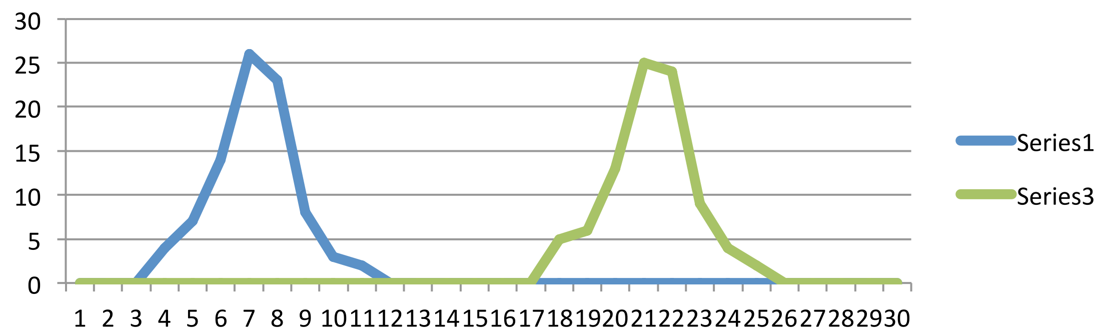
Dynamic Time Warping

- Computes a ‘distance’ between two time series – DTW distance
- Different to Euclidean distance
- The DTW distance can be used as a metric for ‘closeness’ between the two time series

MDSec

Dynamic Time Warping - Example

- Consider the following sequences:
  - 0 0 0 4 7 14 26 23 8 3 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  - 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 5 6 13 25 24 9 4 2 0 0 0 0 0
- Initial analysis suggests they are very different, if comparing from the entry points.
- However there are some similar characteristics:
  - Similar shape
  - Peaks at around 25
  - Could represent the same sequence, but at different time offsets?



MDSec

Side Channel Attacks

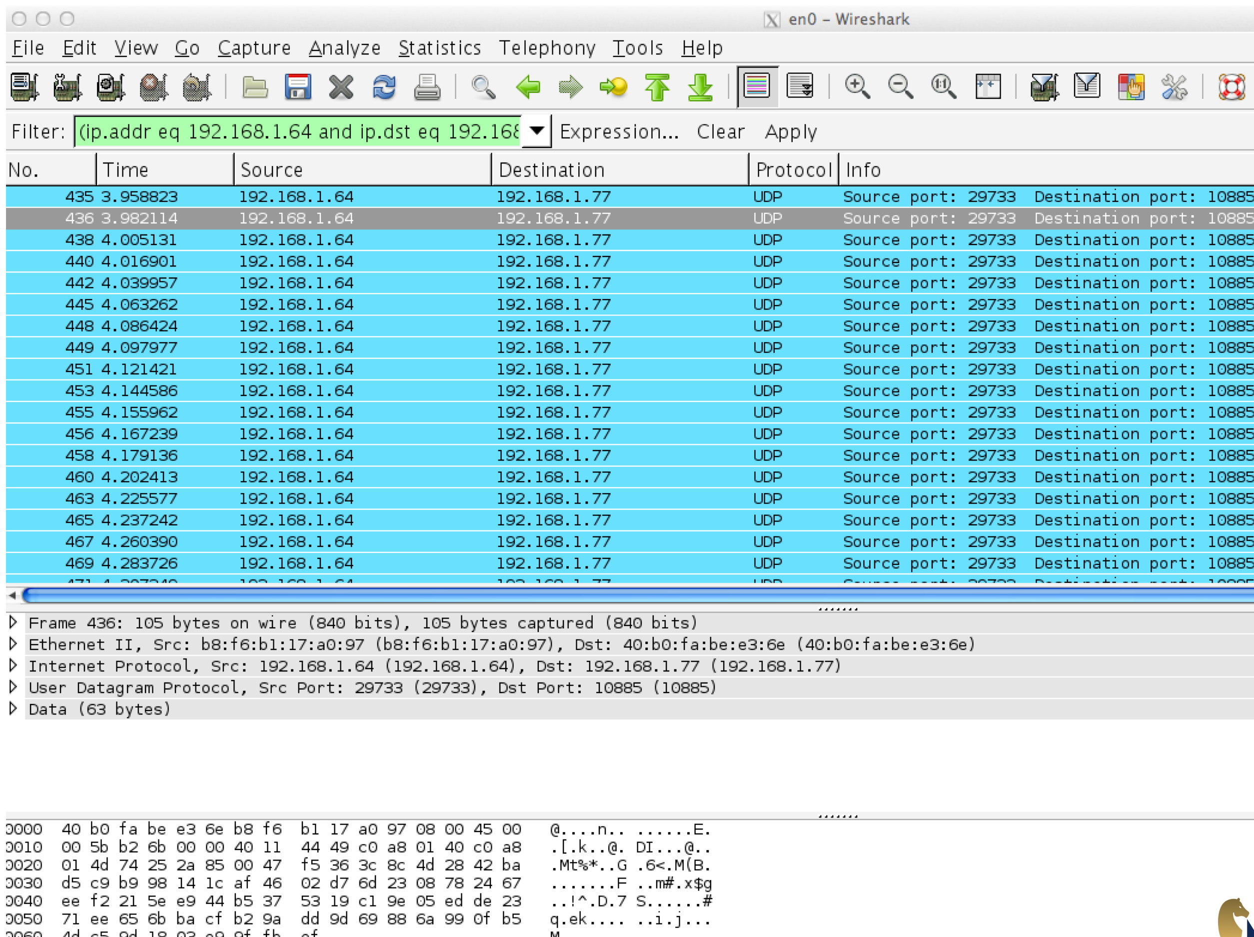
Side Channel Attacks

- Usually connections are peer-to-peer
- We assume that encrypted VoIP traffic can be captured:
  - Man-in-the-middle
  - Passive monitoring
- Not beyond the realms of possibility:
  - "GCHQ taps fibre-optic cables" http://www.theguardian.com/uk/2013/jun/21/gchq-cables-secret-world-communications-nsa
  - "China hijacked Internet traffic" http://www.zdnet.com/china-hijacked-uk-internet-traffic-says-mcafee-3040090910/

MDSec

Side Channel Attacks

- But what can we get from just a packet capture?



MDSec

Side Channel Attacks

- Source and Destination endpoints
  - Educated guess at language being spoken

- Packet lengths

- Timestamps

Packet Lengths with filter: (ip.addr eq 19...

Topic / Item	Count	Rate (ms)	Percent
Packet Lengths	249	0.052311	
0-19	0	0.000000	0.00%
20-39	0	0.000000	0.00%
40-79	1	0.000210	0.40%
80-159	241	0.050630	96.79%
160-319	7	0.001471	2.81%
320-639	0	0.000000	0.00%
640-1279	0	0.000000	0.00%
1280-2559	0	0.000000	0.00%
2560-5119	0	0.000000	0.00%
5120-	0	0.000000	0.00%

Close

Side Channel Attacks

- So what?......
- We now know VBR codecs encode different sounds at variable bit rates
- We now know some VoIP implementations use a length preserving cipher to encrypt voice data

Side Channel Attacks

Variable Bit Rate Codec

+

Length Preserving Cipher =

MDSec

Case Study

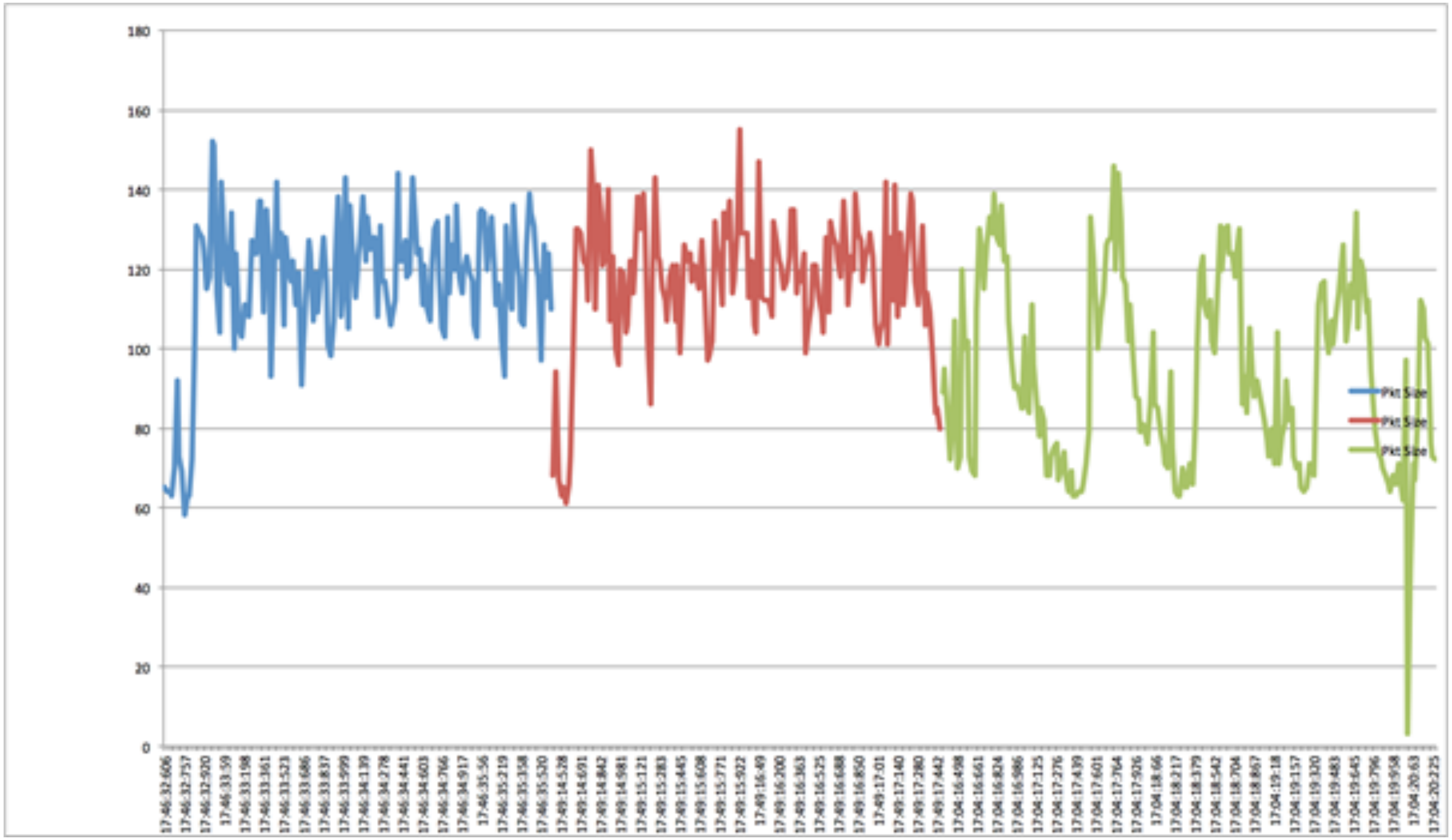
Skype Case Study

- Connections are peer-to-peer
- Uses the Opus codec (RFC 6716):

  "Opus is more efficient when operating with variable bitrate (VBR) which is the default"
- Skype uses AES encryption in integer counter mode
- The resulting packets are not padded up to size boundaries

MDSec

Skype Case Study



Skype Case Study

- Although similar phrases will produce a similar pattern, they won't be identical:
  - Background noise
  - Accents
  - Speed at which they're spoken

- Simple substring matching won't work!

MDSec

Skype Case Study

- The two approaches we chose make use of the NLP techniques:
  - Profile Hidden Markov Models
  - Dynamic Time Warping

Skype Case Study

- Both approaches are similar and can be broken down in the following steps:
  - Train the model for the target phrase
  - Capture the Skype traffic
  - “Ask” the model if it’s likely to contain the target phrase

MDSec

Skype Case Study - Training

- To “train” the model, a lot of test data is required
- We used the TIMIT Corpus data
- Recordings of 630 speakers of eight major dialects of American English
- Each speaker reads a number of “phonetically rich” sentences

Skype Case Study - TIMIT

“Why do we need bigger and better bombs?”

MDSec

Skype Case Study - TIMIT

“He ripped down the cellophane carefully, and laid three dogs on the tin foil.”

Skype Case Study - TIMIT

“That worm a murderer?”

Skype Case Study - Training

- To collect the data we played each of the phrases over a Skype session and logged the packets using tcpdump

```
for((a=0;a<400;a++)); do /
Applications/VLC.app/Contents/MacOS/
VLC --no-repeat -I rc --play-and-exit
$a.rif ; echo "$a " ; sleep 5 ; done
```

MDSec

Skype Case Study - Training

- PCAP file containing ~400 occurrences of the same spoken phrase
- “Silence” must be parsed out and **removed**
- Fairly easy - generally, silence observed to be less than 80 bytes
- Unknown spikes to ~100 during silence phases

Skype Case Study - Silence

Short excerpt of Skype traffic of the same recording captured 3 times, each separated by 5 seconds of silence:

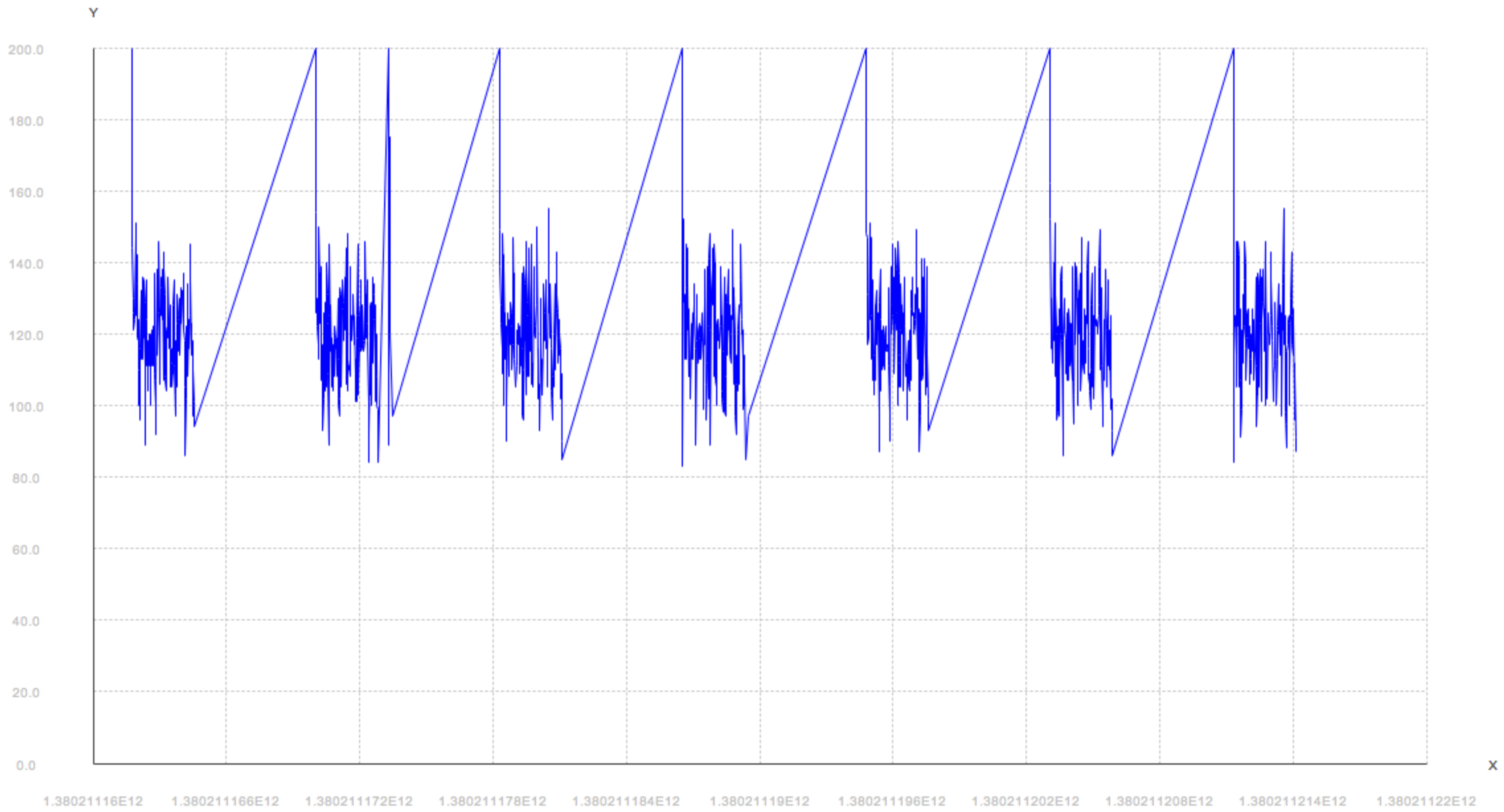
Skype Case Study - Silence

Approach to identify and remove the silence:

- Find sequences of packets below the silence threshold, ~80 bytes
- Ignore spikes when we're in a silence phase (i.e. 20 continuous packets below the silence threshold)
- Delete the silence phase
- Insert a marker to separate the speech phases – integer 222, in our case
- This leaves us with just the speech phases…..

MDSec

Skype Case Study - Silence



MDSec

Skype Case Study – PHMM Attack

- Biojava provides a useful open source framework
  - Classes for Profile HMM modeling
  - BaumWelch for training
  - A dynamic matrix programming class (DP) for calling into Viterbi for sequence analysis on the PHMM

- We chose this library to implement our attack

Skype Case Study – PHMM Attack

- Train the ProfileHMM object using the Baum Welch
- Query Viterbi to calculate a log-odds
- Compare the log-odds score to a threshold
- If above threshold we have a possible match
- If not, the packet sequence was probably not the target phrase

Skype Case Study – DTW Attack

- Same training data as PHMM
- Remove silence phases
- Take a prototypical sequence and calculate DTW distance of all training data from it
- Determine a typical distance threshold
- Calculate DTW distance for test sequence and compare to threshold
- If the distance is within the threshold then likely match

PHMM Demonstration

Skype Case Study – Pre Testing

Skype Case Study – Post Testing

Cypher: “I don’t even see the code. All I see is blonde, brunette, red-head”

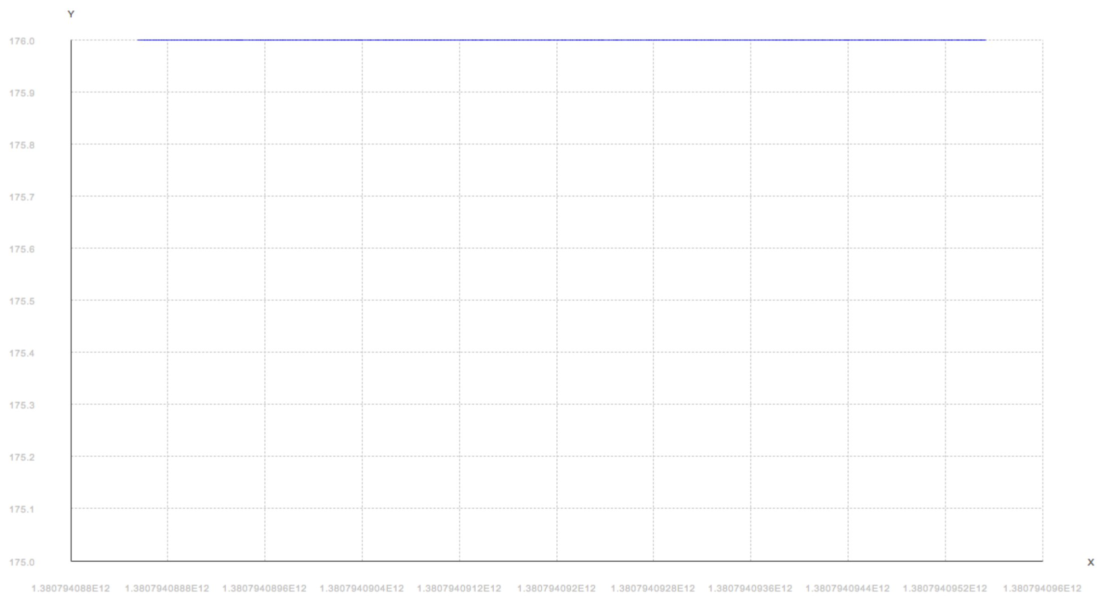
PHMM Statistics

- Recall rate of approximately 80%
- False positive rate of approximately 20%
- Phonetically richer phrases will yield lower false positives
- TIMIT corpus: “Young children should avoid exposure to contagious diseases”

DTW Results

- Similarly to PHMM results, ~80% recall rate
- False positive rate of 20% and under – again, as long as your training data is good.

Silent Circle - Results



- Not vulnerable – all data payload lengths are 176 bytes in length!

Wrapping up

Prevention

- Some guidance in RFC656216
- Padding the RTP payload can provide a reduction in information leakage
- Constant bitrate codecs should be negotiated during session initiation

Further work

- Assess other implementations
  - Google Talk
  - Microsoft Lync
  - Avaya VoIP phones
  - Cisco VoIP phones
  - Apple FaceTime
    - According to Wikipedia, uses RTP and SRTP…Vulnerable?
- Improvements to the algorithms - Apply the Kalman filter?

Conclusions

- Variable bitrate codecs are unsafe for sensitive VoIP transmission
- It is possible to deduce spoken conversations in encrypted VoIP
- VBR with length preserving encrypted transports like SRTP should be avoided
- Constant bitrate codecs should be used where possible

@domchell @MDSecLabs