CERIAS

The Center for Education and Research in Information Assurance and Security

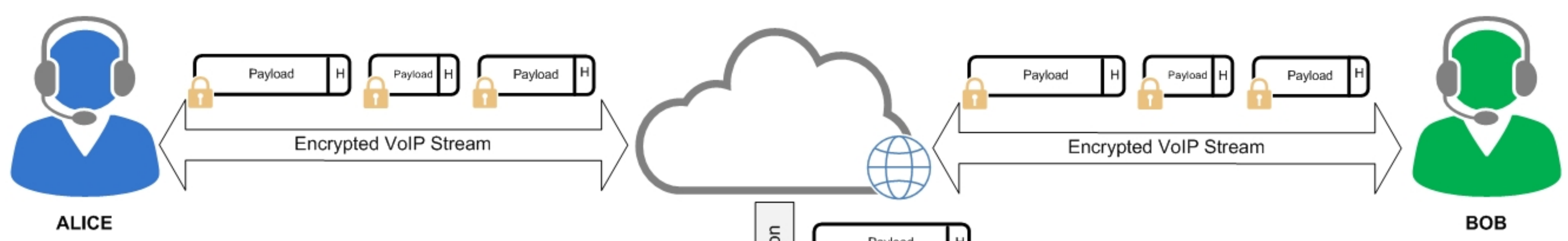
Random Packet Size Morphing and Direct Target Sampling for VoIP Language Obfuscation

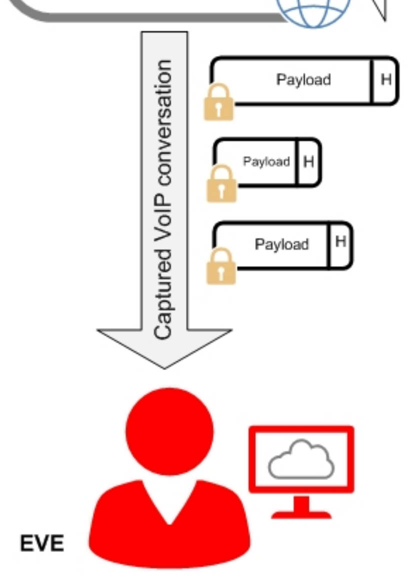
Veronica Manfredi[a] , Filipo Sharevski[b], Melissa Dark, PhD[c]

1. Packet Size Attacks on Encrypted VoIP

Background: QoS and VBR

- QoS: minimal delay, negligible jitter, no packet loss
- Practical implementation: VBR → CELP → Speex[1]
- VBR codecs encode easier sounds at lower bit rates → better QoS





Security measures

- ✓ Encrypted payload (SRTP and SRTCP)
- ✓ Length-preserving stream cipher

Packet size attacks may reveal:

- ✓ Spoken language
- ✓ Key phrases
- ✓ A transcript of the conversation

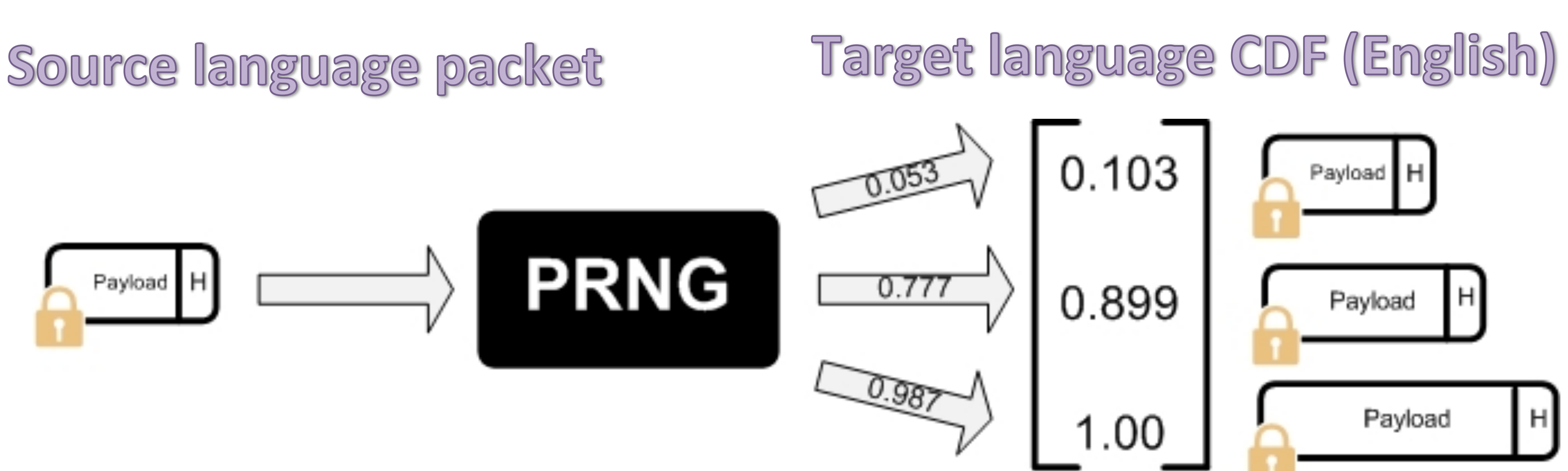
Security flaws

- X Packet sizes correspond to spoken sounds
- X Individual or sequential (**n-gram**) packet sizes leak information

2. Language Identification Countermeasures

Direct Target Sampling (DTS)[2]

- Pads and splits packets to match the distribution of n-grams in a target language
- Reduces the language identification classifier's accuracy
- Adds a significant amount of padding, increasing overhead



Traffic Morphing (TM)[3]

- Uses convex optimization techniques to solve the equation Y = AX
  - Y is the target language n-gram distribution vector
  - X is the source language n-gram distribution vector
  - A is a morphing matrix obtained by solving an optimization problem that minimizes the difference between source and target packet size

3. Random DTS and TM

DTS and TM countermeasures fail to reduce classifier accuracy when the classifier is:

- able to train on examples of morphed traffic, and
- at least an order higher than the applied countermeasure[3]

Random DTS and TM generate a random target distribution for each VoIP conversation to remove the ability to train on morphed traffic and reduce the classifier's accuracy.

4. Experimental Design

- English, Hungarian, Mandarin, Portuguese, Russian, Tamil languages[4] → 9 packet sizes produced by Speex codec
- Binary X^2 **trigram** morphing and DTS **trained classifier**: 8 characteristic trigrams, 93 speakers per language
- Random DTS and TM on **unigrams**
- CVXOPT solver interface for the following optimization problem[3]:

Minimize: $\sum_{\forall i,j\in[1,n]} x_j a_{ij}(|s_i - s_j|)$

Subject to: $\sum_{j=1}^{n} a_{ij} x_j = y_i \quad \forall i \in [1,n],$

$\sum_{i=1}^{n} a_{ij} = 1 \quad \forall j \in [1,n],$

$a_{ij} \geq 0 \quad \forall i,j \in [1,n]$

x_j : source packet

y_i : target packet

s_i : source packet size

s_j : target packet size

$a_{i,j}$: probability of morphing s_j to s_i

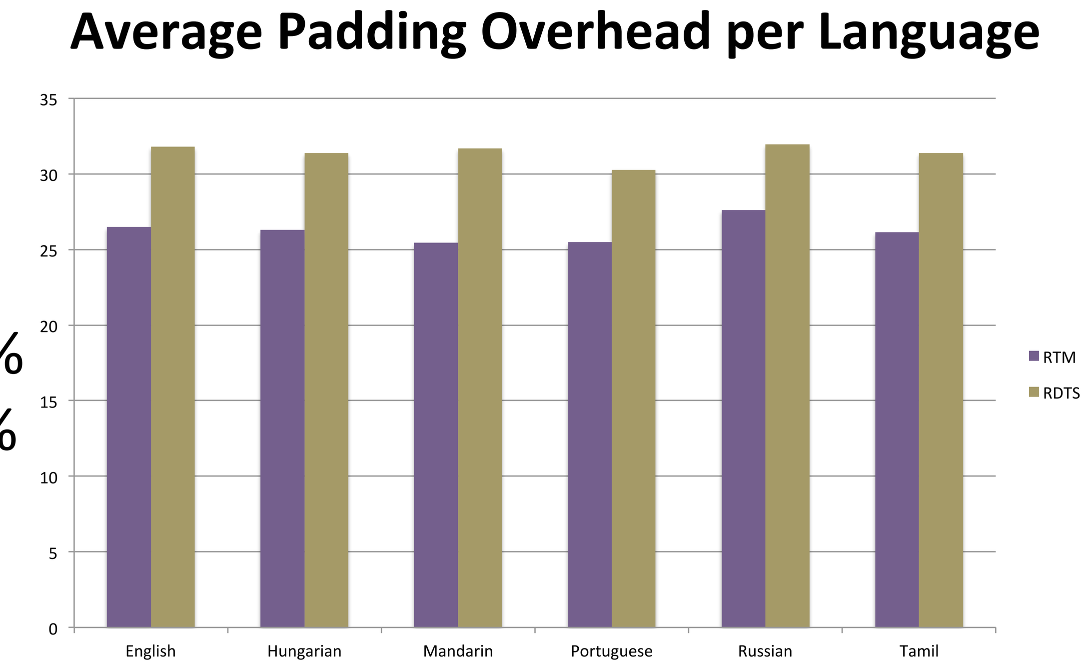
5. Results

Average padding overhead

Random DTS: 31.4%

Random Morphing: 26.2%

Average Padding Overhead per Language



Accuracy Effects on Strongest Cases

Countermeasure	Russian v. English	Russian v. Portuguese	Tamil v. English	Mandarin v. English	Hungarian v. Portuguese
none	0.886	0.818	0.818	0.795	0.795
Random TM	0.500	0.568	0.386	0.568	0.477
Random DTS	0.500	0.614	0.409	0.614	0.500

Average classifier accuracy

Original sizes: 69.8%

Random Morphing: 51.2%

Random DTS: 50.6%

Average classifier accuracy reduction

Random Morphing: 25.1%

Random DTS: 26.5%

6. Conclusion

Random DTS and TM reduced language-related information leakage, but add padding overhead. More work is needed to examine their effectiveness against phrase detection, transcript reconstruction, timing attacks, and other classifiers, as well as their impact on QoS.

a. Bunker Hill Community College, Computer Information Technology, Boston, MA.

b., c. Purdue University, CERIAS, Department of Computer and Information Technology, West Lafayette, IN.

1. Speex Voice Codec. http://www.speex.org/

2. Dyer, K. P., Coull, S. E., Ristenpart, T., & Shrimpton, T. (2012). Peek-a-boo, I still see you: Why efficient traffic analysis countermeasures fail. Paper presented at the 2012 IEEE Symposium on Security and Privacy (SP).

3. Wright, C. V., S. E. Coull, and F. Monrose, 2009, Traffic morphing: An efficient defense against statistical traffic analysis: Proceedings of the Network and Distributed Security Symposium-NDSS.

4. *Center for Spoken Language Understanding (CSLU), Oregon Health and Science University. 22 Language V1.2.* http://www.ohsu.edu/cslu

This material is based upon work supported by the National Science Foundation under grant #1062970. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

CERIAS®

PURDUE UNIVERSITY