The Joys (and Challenges) of Inter- & Cross-Disciplinary Security Research

Fabian Monrose
About me

- Professor of Computer Science, UNC Chapel Hill
- Broad interests in computer and communications security
UNC Chapel Hill

- Founded in **1964** by Prof. Fred Brooks
- Second oldest Ph.D. granting CS department in the U.S.
  - M.S. and Ph.D program (since 1965)
    - currently ~160 students
  - B.S. program (since 2001)
    - currently ~300 majors
The Joys and Challenges of Cross-Disciplinary (Security) Research

Research areas

- Algorithms and complexity theory
- Assistive technology
- Bioinformatics and computational biology
- Computer graphics
  - **Computer security**
- Computer-supported cooperative work
  - **Computer vision**
- Databases and data mining
- Distributed systems
- Geometric modeling and computation
  - **Hardware systems**
- High performance computing
- Medical image processing
- Multimedia
  - **Networking**
- Physically-based simulation
- Real-time systems
- Robotics
- Software engineering and environments
  - **Theorem proving and term rewriting**
A great place!

- Great research and learning opportunities
- Great facilities
- Great working environment
  - congenial, collegial, collaborative and exciting
- I’m always on the lookout for talented **students** and **postdocs**!
In Pasteur’s Quadrant: Innovation and Research @UNC

- desire to solve practical problems
- pragmatic engineering and use-inspired basic research

Recent interests in PQ

- Traffic analysis of encrypted communications:
  - “opaque” traffic analysis [NDSS’13]; New mitigation strategies (VoIP)

- Network Security:
  - detecting & mitigating code injection attacks [USENIX Sec’11]
  - fast multi-dimensional querying of compressed network payloads [USENIX ATC’12]
  - detecting network malfeasance via sequential hypothesis tests [DSN’13]
  - fast and efficient subtree mining for situational awareness

- Computer Forensics:
  - tracking and mediating accessing to digital objects [CCS’11]
  - deploying secure virtual data enclaves [DHS;RENCI]
Recent interests in PQ

- Operating Systems Security:
  - New frameworks: just-in-Time Code Reuse [S&P’13];
  - Revisiting fine-grained ASLR [USENIX Sec’14]
  - Limiting Memory Exploits [NDSS’15]
  - New OS-level protection primitives [major focus of my ongoing work]

- Computer visions meets security
  - Compromising reflections [CCS’13,’14]
  - Enabling privacy-preserving situational awareness from massive video collections [ongoing work]
  - Content-based copy detection
Today’s talk

Examine some interesting (at least to me) problems requiring multidisciplinary teams to solve them

- focus on the **journey**, rather than the end result
- highlight challenges (and open problems) common across experiences:
  - linguistics and computer security
  - computer vision and computer security
  - machine learning and machine translation

- focus on traffic analysis of encrypted VoIP communications
Voice over IP (VoIP)

- Popular replacement for traditional telephony
- Many free, or inexpensive, services available
  - very reliable
  - easy to use
VoIP Security

- Security and privacy implications are still not well understood
- Two channels: **voice** and **control**
- Majority of security analyses focus on control channel
  - e.g., caller id spoofing, registration hijacking, denial of service

We are interested in the **secrecy** of the **voice** channel
Information leakage

Voice channel is encrypted to ensure confidentiality

<table>
<thead>
<tr>
<th>codec</th>
<th>SRTP</th>
<th>UDP</th>
<th>IP</th>
</tr>
</thead>
</table>

1 frame (~20ms audio) per packet
Information leakage

Two important design decisions:

**compression**: variable-bit-rate (VBR) codecs
- compress different sounds with varying fidelity

**encryption**: length-preserving stream ciphers
An unintended interaction

result: packet sizes reflect properties of the input signal
Privacy on the line

- examine feasibility of eavesdropping after encrypted calls have been established

- we do NOT learn the encryption key
- instead apply traffic analysis to decode a stream
- K. McCurley’s work “Language Modeling and Encryption on Packet Switched Networks”, noted similar problems in 2006

Credit: W. Diffie and S. Landau
Where is this combination supported?

*uses speex codec
Finding the right talent

- Charles Wright @PSU
- Lucas Ballard @Google
- Scott Coull @Redjack
- Andrew White (@UNC)
- Austin Matthews @CMU
- Kevin Z. Snow (@UNC)
How bad is this leak?

Sufficient to determine:

2007
- Wright et al.; Language identification of encrypted VoIP traffic: *Alejandra y Roberto or Alice and Bob?*, USENIX Security

2008
- Wright et al., *Spot me if you can: Uncovering spoken phrases* in encrypted VoIP conversations, IEEE S&P

2009

Prior work did *not* take advantage of language-specific constraints or permitted sequences (i.e., “phonotactics”)
Finding the right talent

- We have little training in linguistics. What do we do?
  - outsource or recruit talent beyond Computer Science?
  - educate ourselves --- long hours in the library and classes?
  - quit and move on?
Finding the right talent

Andrew White (@UNC)
Austin Matthews (@CMU)
Kevin Z. Snow (@UNC)
Elliott Moreton, Roger Que (@UNC)
Katherine Shaw (@UNC)
Phonetic Models of Speech

- represent speech as a sequence of **phonemes**
  - individual speech units
  - based on articulatory processes
  - airflow through the mouth, throat and nose
- about 50-60 phonemes in **English**
- representation: International Phonetic Alphabet (**IPA**)
The Joys and Challenges of Cross-Disciplinary (Security) Research

Rock and roll!

VoIP Conversation Reconstruction Process

Encrypted VoIP Packets

Packet Lengths

Phoneme Segments

Phonemes

Corrected Phonemes

Word Segments

Words

rock

and

roll
Infants use perceptual, social, and linguistic cues to segment the stream of sounds

- use learned knowledge of well-formedness
  - amazingly, infants learn these rudimentary constraints while simultaneously segmenting words
- use familiar words (e.g., their own name, “mama,” etc) to identify new words in a stream


• As early as 6 months, **within-word** versus **between-word sounds learned**

Step 1: phonetic segmentation

Rock and roll!

Encrypted VoIP Packets → Packet Sizes → Phoneme Segments
Step 1: phonetic segmentation

IPA Pronunciation of the phrase “an official deadline”

Observation: frame sizes tend to differ in response to phoneme transitions
Segmentation: Approach

- identify **boundaries** using a **discriminative** machine learning technique that takes contextual features, and our history of decisions, into consideration

- model only those features which help distinguish between classes

- we apply concepts similar to those proposed by Hayes and Wilson, 2008.

Step 2: Phoneme classification

Rock and roll!

Encrypted VoIP Packets

Phoneme Segments

Phonemes

Corrected Phonemes

The Joys and Challenges of Cross-Disciplinary (Security) Research

Fabian Monrose
Step 2: phoneme classification

Observation: differing sounds are **encoded** at different bit rates (e.g., **Speex** codec only uses **9 different bit rates** in narrow band mode; 21 bit rates in wide-band mode)
Phoneme classification: Features

- nearby sequences
- popular bigram / trigrams

- We also apply language model correction
  - incorporates contextual information
  - “corrects” misclassification
Step 3: Word break insertion

Rock and roll!

Encrypted VoIP Packets

Corrected Phonemes

Word Segments

- Encrypted VoIP Packets
- Corrected Phonemes
- Word Segments
Step 3: Word break insertion

Based on language-specific constraints on **phoneme order**

- Insert potential word breaks into *impossible* phonetic triplets
  - [ɪŋw] (‘blessing way’)
- Resolve **invalid** word beginning / endings
  - [zdr] (‘eavesdrop’)
- Improvement: split resulting segments by **dictionary search**

Word Breaks: Method

- find all dictionary word sequences where their pronunciation matches our segment
- insert consistent word breaks

<table>
<thead>
<tr>
<th>input:</th>
<th>ðıIksaItIŋweI</th>
</tr>
</thead>
<tbody>
<tr>
<td>dictionary sequences:</td>
<td>ðı•IksaItIŋ•weI  'the exciting way'</td>
</tr>
<tr>
<td></td>
<td>ðı•Ik•saItIŋ•weI  'the ick citing way'</td>
</tr>
<tr>
<td>consistent breaks:</td>
<td>ðı•IksaItIŋ•weI  'the exciting way'</td>
</tr>
</tbody>
</table>
Stage 4: Word Matching

Encrypted VoIP Packets

\[ \text{rock} \quad \text{and} \quad \text{roll} \]

Word Segments

We are computer scientists! We can do this!
Word Matching: Method

Find closest pronunciation using an edit distance approach to infer articulatory distance between phonemes.
Vowels

- characterized by tongue position and lip shape
- **height**: height of the tongue
- **backness**: tip of the tongue forward or backward
  - (e.g., ‘there’, ‘here’)
- **rounding**: lip pucker
Consonants

• characterized by restriction of airflow

  **place:** *where* the restriction is made

  **manner:** *how* the restriction is made

• also by **voicing:** whether the vocal chords vibrate

  • several classes, e.g., stops and fricatives
Word Matching: Method

Problem: homophones
(e.g., “ate”, “eight”)

Solution: language model correction using trigram over both words and part of speech tagging
20 months later:

Voice over IP (VoIP) Conversation Reconstruction Process:

- Encrypted VoIP Packets
- Packet Lengths
- Phoneme Segments
- Phonemes
- Corrected Phonemes
- Word Segments
- Words

Phoneme Segmentation
Phoneme Classification
Language Model Correction
Word Segmentation
Word Classification

Rock and roll!
Evaluation

- TIMIT Corpus: 630 speakers (70% male) 10 sentences per speaker
- 8 major dialects of American English

**time-aligned word and phoneme transcripts**
Challenge: How do we evaluate our guesses?

<table>
<thead>
<tr>
<th>reference</th>
<th>It’s not easy to create illuminating examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>hypothesis</td>
<td>Is not except to create illuminated examples</td>
</tr>
</tbody>
</table>

- Automated evaluations for machine translations
- *active* research area (many open problems)
- **METEOR**: Metric for Evaluation of Translation with Explicit ORdering by Lavie and Denkowski, 2007
The METEOR Metric

**Adequacy**: scores the proportion of matching words in the reference (recall) to the proportion of matching words in the hypothesis (precision)

**Fluency**: penalizes fragmentation by matching contiguous subsequences

METEOR Score Interpretation (Lavie, 2010)
Examples

- **METEOR Score Interpretation (Lavie, 2010)**

<table>
<thead>
<tr>
<th>Example 1</th>
<th>Example 2</th>
<th>Example 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>that is your headache</td>
<td>is it's not easy to create illuminated examples</td>
<td>cliff was soothe by a luxurious massage</td>
</tr>
<tr>
<td>METEOR Score: 0.18</td>
<td>METEOR Score: 0.53</td>
<td>METEOR Score: 0.78</td>
</tr>
</tbody>
</table>

UNDERSTANDABLE | GOOD/FLUENT

.1 .2 .3 .4 .5 .6 .7 .8 .9
Hypotheses
(context dependent results)

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SA1</strong>: “She had your dark suit in greasy wash water all year”</td>
<td></td>
</tr>
<tr>
<td>She had year dark suit a greasy wash water all year</td>
<td>0.67</td>
</tr>
<tr>
<td>She had a dark suit a greasy wash water all year</td>
<td>0.67</td>
</tr>
<tr>
<td>She had a dark suit and greasy wash water all year</td>
<td>0.67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SA2</strong>: “Don’t ask me to carry an oily rag like that”</td>
<td></td>
</tr>
<tr>
<td>Don’t asked me to carry an oily rag like that</td>
<td>0.98</td>
</tr>
<tr>
<td>Don’t ask me to carry an oily rag like dark</td>
<td>0.82</td>
</tr>
<tr>
<td>Don’t asked me to carry and oily rag like dark</td>
<td>0.8</td>
</tr>
</tbody>
</table>

SA2: “Don’t ask me to carry an oily rag like that.”
Hypotheses
(context independent results)

<table>
<thead>
<tr>
<th>Reference Hypothesis</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change involves the displacement of form. Codes involves the displacement of aim.</td>
<td>0.57</td>
</tr>
<tr>
<td>Artificial intelligence is for real. Artificial intelligence is carry all.</td>
<td>0.49</td>
</tr>
<tr>
<td>Bitter unreasoning jealousy. Bitter unreasoning dignity.</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Context independent results (New England dialect)
Challenges

• None of the translation scoring techniques (we are aware of) really model how well the translation captures the essence of the conversation

• Techniques for evaluating the adequacy of generated text (be that a translation, automated captions, dialogues, etc) are desperately needed in many subfields of computer science

• many subfields can benefit from advancements in this space
Mitigation

- pad up to multiples of the block size?
- use constant bit rate codecs?
- dynamically insert **multiple frames** per packet?

- IETF (RFC 6366; Aug. 2013) discusses requirements for new Internet Audio Codecs

☑️ Calls for new approaches; several teams are exploring new designs (great area of research)

☐ Challenge: designing **automated** mechanisms for measuring **perceived call quality**
Summary (VoIP)

- VoIP is here to stay. But, security and privacy issues should not be overlooked
  - quality of reconstructed transcripts better than expected
  - will improve with advancements in computational linguistics

We need stronger, **interdisciplinary**, partnerships in order to push the boundaries and design more secure and efficient solutions

Finding the right talent

• As computer scientists we use rapid prototyping as a feedback mechanism

  • this is rare in other disciplines; avoid programming at all costs :)

• Very different expectations with respect to evaluations

  • collaborators in Linguistics expect lab-based evaluations

  • generally not sufficient for our work, and particularly difficult in this case (e.g., for cross-fold evaluations)

• In many cases, we don’t speak a common language

  • but, that’s not a bad thing at all.
Finding the right talent

• Our training in adversarial thinking can be a **curse:**
  - too quick to think of why something **won’t work** (and how to BREAK it!)

• and a **blessing:** we tend to enjoy thinking of ways to push boundaries

Closing Remarks (to students)

• Successful collaborations can help push boundaries in ways you may not have thought possible; and even rethink old problems

• Cross disciplinary research can be highly rewarding

• We need to overcome “cultural practices/biases”, be open to realizing how much we really don’t know and look outside CS.
  
  • fully expect that progress will be slower than you anticipate

• But most of all, have fun!
The Joys and Challenges of Cross-Disciplinary (Security) Research

Thanks!

Fabian@cs.unc.edu