Vuvuzela

Jelle van den Hooff, David Lazar, Matei Zaharia, and Nickolai Zeldovich

~ MIT CSAIL ~

Presentation by: Alexander Gaidis
Goal?

Prevent an adversary from learning information about a single individual

How?

Identify and minimize the number of observable variables and then obfuscate these with differential privacy.
Pathway to Metadata Private Communication

Problem Setup
Tor is in style! Let’s use it!

But what is Tor?
Pathway to Metadata Private Communication

Attempt #1: Single Server
Pathway to Metadata Private Communication

Attempt #2: Dead Drops

Dead Drop: 792F423F4428472B4B6250655368566D
Message: "Trump’s tax returns are…"

Dead Drop: 792F423F4428472B4B6250655368566D
Message: ""

Dead Drop: 2B4D6251655468576D5A713474377721
Message: ""

ROUND 1

(The Server)
Pathway to Metadata Private Communication

Attempt #2: Dead Drops

The Intercept

Dead Drop: 792F423F4428472B4B6250655368566D
Message: H+MbQeThWmZq4t7w9z$C&F)J@NcRfUjX

Dead Drop: 792F423F4428472B4B6250655368566D
Message: VmYq3t6w9z$C&E)H@McQfTjWnZr4u7x!

Dead Drop: 2B4D6251655468576D5A713474377721
Message: w!z%C*F-JaNdRgUjXn2r5u8x/A?D(G+K

(The Server)

ROUND 1
Pathway to Metadata Private Communication

Attempt #2: Dead Drops

ROUND 1
Pathway to Metadata Private Communication

Attempt #3: Mixnet

But what’s a mixnet?

(The Last Server)
Pathway to Metadata Private Communication

Attempt #4: Noise
Let $d_2 = \# \text{ dead drops with two accesses in a single round. Then,}$

$$\Pr[d_2 = x \mid Alice \text{ talked to Bob}] \approx \Pr[d_2 = x \mid Alice \text{ did not talk to Bob}]$$
We achieve differential privacy through the addition of noise.

\[
\Pr[d_2 = x \mid Alice\ talked\ to\ Bob] \approx \Pr[d_2 = x \mid Alice\ did\ not\ talk\ to\ Bob]
\]
Scenario:
- Assume Eve is Evil
- Alice talks to Eve through Vuvuzela
- The NSA arrests Alice for being an accomplice to Eve
- Will a jury convict Alice?
Walkthrough of the Full System
Outline

- Pathway to Metadata Anonymous Communication
- Dialing Protocol
- Threat Model Recap & Analysis
- Results
- Discussion
# The Last Piece: The Dialing Protocol

<table>
<thead>
<tr>
<th>Communication Protocol</th>
<th>Dialing Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversation Dead Drops</td>
<td>Invitation Dead Drops (much larger)</td>
</tr>
<tr>
<td>Conversation Round &lt; 1 Minute</td>
<td>Dialing Round = 10 Minutes</td>
</tr>
<tr>
<td>1 Message = 240 Bytes</td>
<td>Invitation Download = Variable Size</td>
</tr>
<tr>
<td>Responses Travel through Mixnet</td>
<td>Invitations Downloaded Directly</td>
</tr>
</tbody>
</table>
Outline

Pathway to Metadata Anonymous Communication

Dialing Protocol

Threat Model Recap & Analysis

Results

Discussion
Threat Model

• N – 1 Servers Compromised
• Complete Network Surveillance
• X Sybil Clients
• Interference over Multiple Rounds

Trust Model

• 1 Honest Server
• You and Your Friend are Honest
• Honest Client/Server Runs Bug-Free Code
Outline

Pathway to Metadata Anonymous Communication

Dialing Protocol

Threat Model Recap

Results

Discussion
Results

Figure 9: Performance of Vuvuzela’s conversation protocol when varying the number of users online. Every user sends a message every round.
Figure 11: Performance of Vuvuzela’s conversation protocol when varying the number of servers with 1 million active users and $\mu=300,000$. 
### Results

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant bandwidth cost for client (in convo protocol)</td>
<td>Dialing protocol is expensive for clients</td>
</tr>
<tr>
<td>Protection against a strong adversary</td>
<td>Dialing protocol is not forwardly secret</td>
</tr>
<tr>
<td>System can be abstracted, leaving a clean messaging UI</td>
<td>Bandwidth cost incurred by servers</td>
</tr>
<tr>
<td>More users = more traffic = more privacy</td>
<td>Does not guarantee group privacy</td>
</tr>
<tr>
<td>Security guarantee holds with many or few users</td>
<td>Sending tons of messages degrades privacy</td>
</tr>
<tr>
<td></td>
<td>Fixed message size, roughly as big as a tweet</td>
</tr>
<tr>
<td></td>
<td>Infrequent dialing rounds</td>
</tr>
</tbody>
</table>
You Asked, We Will Try to Answer

Group chat?

Improve scalability while maintaining privacy?

Can the dialing and conversation protocols happen at the same time?

If dead drops are erased each round, how does retransmission work?

Would adding random delays to messages (stall message passing) give the same guarantees as shuffling messages?