CS144r/244r
Network Design Project
on
Secure and Intelligent Internet of Things

(iot security 3: solution)

3/10/2014

Instructor: Professor HT Kung
Harvard School of Engineering and Applied Sciences
Announcements

• Tomorrow student teams will individually receive feedback on their 1-page abstracts of proposed course projects submitted last night.
• We will have no quiz problems this week, so students should focus on refining their project plans. Students may send in updated versions by midnight tonight.
• On Wed (March 12), Dr. Trevor Blackwell will give a guest lecture on “Commercializing Connected Devices: Experience from Y Combinator's hardware startups”.
• Project proposal presentations will be on March 24 and 26 (class will be from 2:30-4:30pm on these two dates). See the updated course schedule on the course web site.
• On April 9, Dr. Robert Cohn will give a second lecture on IoT programming, which will give programming advices concerning student course projects.
• Please start thinking about what research papers your team will present. Link to the paper needs to be submitted for approval by Tuesday midnight, March 25.
• See updated course schedule on the course website.
• For those teams who need parts to be ordered, please let TFs know.
• The Air Force Research Laboratory in NY and Intel-NTU Center in Taipei are very interested in summer interns in IoT.
What Is New in the News?

1. “Internet Of Things, Data To Change Online Advertising Forever” (3/9/2014)… Internet-connected devices will change online advertising in ways that industry executives cannot yet imagine. …

The picture on the right illustrates mobile ads concerning items related to the fridge

Automotive manufacturers are among the first to build a business model -- connected cars – for IoT online advertising. Agencies can now push information to the consumer, telling them to take a specific route to save money and gas. And, oh, by the way, stop at the Chevron station at Broadway and Third to fill up because they have the least expensive gas in the area

In each case, consumer behavior is driven less by brand loyalty (though that plays a part) than convenience. If you're a savvy marketer, you can catch your target consumer right before they buy something. Why spend millions on TV advertising when you can do that?

2. “Internet of things’ funding pledge” (3/9/2014)…UK prime minister, David Cameron, has announced that the UK government will spend an extra £45m on developing so-called “internet of things” technology, speaking at the CeBIT technology trade fair this week in Germany

Mr Cameron said the UK and Germany could find themselves on the forefront of a new “industrial revolution”. “I see the internet of things as a huge transformative development – a way of boosting productivity, of keeping us healthier, making transport more efficient, reducing energy needs, tackling climate change”
Recap: IoT Networking

• Long-term vision
  – All Things (sensors/actuators with microcontrollers) will connect to the Internet using Internet Protocols (IP), perhaps via gateways
  – Existing sensors will be joined by many, many, millions more, so that anything worth measuring can be measured ... and then managed

• For such scaling, we must re-think about IP based networking

• We discussed the current thinking of IETF in this area under the IPv6 framework, i.e., 6LoWPAN for constrained networks and RPL for their routing. (Note that IPv6 addresses are 128 bits long!)
Recap: IoT Security (Basic Concepts)

• Security is about protection of information
  – The CIA triad: Confidentiality, Integrity and Availability

• Privacy is about governance and use of information
  – Security is necessary for privacy, but is not sufficient (the hospital example)
Recap: Key Management

• Security schemes which require all parties to share a secret could be difficult to implement, especially for environments under distributed management

• We consider three classical examples which can work without such requirements
  1. **Diffie-Hellman** key agreement protocol (e.g., establishment of symmetric key)
  2. **Hash chain** (e.g., one-time password)
  3. **Asymmetric key** system (e.g., public key and certificate)

• All these methods depend on the fact that some “inverse” function is hard to compute. E.g.,
  - It is easy to compute \( u = p \cdot q \) from \( p \) and \( q \), but it is hard to compute \( p \) and \( q \) from \( u = p \cdot q \) (factorization is hard)
  - For a given \( g \), it is easy to compute \( v = g^a \) from \( a \), but it is hard to compute \( a \) from \( v = g^a \) (logarithm is hard)
Recap: Three Types of IoT Attacks

1. Malware exploiting bugs or vulnerabilities with malicious code
   - **Buffer overflow** and attacks on SD cards

2. Exploitation of lax security policies/practice
   - **ThingBot** and attacks on smart light bulb systems, smart cameras, etc.

3. Link-layer attacks
   - Attacks on **Bluetooth Smart** links, implantable and wearable medical devices, etc.
Today’s Topic: IoT Security Solution

● Focus on new innovative solution approaches for IoT consumer market

● In particular, we will discuss two approaches
  ● Use of software defined networking in helping manage security for IoT devices
  ● Use biometric signature for user authentication: gait recognition and ECG recognition
Conventional Security Practice for Networked PCs

● Device Level
  ● **Secure booting**. Only signed software will be loaded
  ● **Access control**. The principle of least privilege is used. That is, you only have access to what you need to have
  ● **Device authentication**. Devices authenticate themselves using configured or built-in keys
  ● **Firewalls, and intrusion detection and prevention systems** (IDS/IPS), with black and white listing

● Network level
  ● VPN, physical media encryption
Challenges in IoT Security: A Summary

- Security measures cannot assume heavy involvements from users, nor IT support
- Often security considerations have serious implications in safety
- Constrained computing and communication resources (consider, e.g., routers at home)
  - Can’t filter common Internet traffic
  - Difficult for IoT devices to authenticate themselves, receive security patches, and install software update
Some Principles of IoT Security Solution

1. Recall that the power of IoT stems from sensing, networking and self-learning. We will leverage the same technologies in constructing security solution.

2. Use separate security approaches for enterprise and consumer IoT markets.
   - For enterprise market, we can leverage conventional security measures like those outlined in the preceding slide and those designed for deeply embedded endpoints in factory floor automation.
   - For consumer market, we will need new frameworks and technologies, such as software defined networking (SDN) and biometric authentication to be presented next. The latter mitigates the difficulty of involving active interactions from humans in managing IoT security.
Software Defined Networking (SDN): Introduction

• SDN is a new approach to network design and management
• SDN decouples the:
  – Network control plane (systems that make decisions about how to route traffic)
  – Data plane (systems that actually route data packets)
• OpenFlow is the most popular SDN standard protocol enabling interaction between the two planes (although there are others)

For more introductory information see http://www.zdnet.com/10-key-questions-about-software-defined-networking-sdn-7000015822/
Benefits of SDN

• Directly programmable
  – Because the control plane is separate from the data plane, the network can be programed

• Agile
  – Network-wide traffic can be dynamically adjusted

• Centrally managed
  – Administrators get a centralized view of the network to allow them to manage the entire network

• Programmatically configurable
  – Networks can be configured programmatically, allowing for rapid deployment and changes to the network
SDN and IoT

• Security Features Added by SDN
  – Automatically quarantine infected IoT devices
  – Network Slice Virtualization to keep different classes of devices on different networks (reducing the ability of an IoT device to serve as attack vector to another IoT device)
  – Route traffic through next-gen firewall and intrusion detection system
  – Reroute traffic dynamically to restore service to network under DDOS attack

• Example SDN Use Case
  – TV is connected via network for content and updates
  – Attackers can exploit vulnerability in the TV in a corporate network, and then use it to stage an attack against a sensitive computer
  – SDN allows network administrators to quarantine infected IoT devices, so the TV can be quarantined if exhibiting odd behavior or infected, reducing the risk of an attacker using it to access other devices on the network

Source: McAfee IoT and SDN 2014 Threat Predictions
Unobtrusive Security in IoT

• Traditional security measures use key management systems and challenge-response mechanisms
  – As we noted, these schemes can be a poor fit for IoT
    • Devices have limited input for challenge-response leading to tedious mechanisms (Xbox controller for keyboard input…)
    • Key management is expensive and RNGs on small devices can be weak (Bluetooth LE)

• Need mechanisms that provide authentication w/o placing large burden on users

• Possible concept to consider:
  – Passive authentication for soft security (such granular security is often sufficient for IoT scenarios such as identifying a user in a small set of family members in a home)
Passive Authentication

• IoT provides immense wealth of sensors (ubiquitous)
  – Accelerometers, gyroscopes, magnetometers, cameras,
    depth-cameras, microphones, etc.
• Each of these can listen/sense without active user
  interaction
• What are routine behaviors that humans perform that
  may be unique enough for recognition?
  – **Gait** (walking – motion sensors)
  – Speech/Conversation
  – Location/Environment
• Let’s use Gait Recognition as example....
Gait Signal in Time-series

- Measurements taken over time is referred to as \textbf{time-series} data.
- We can use time-series signals generated by an individual’s gait to uniquely identify the individual.
  - Relevant sensors include camera, accelerometers, gyroscopes.
- Such MEMS sensors are in smart phones, smart watches and headsets and very power efficient.

Accelerometer Time-series

Gyroscope Time-series
Time-series Analysis

• Speech recognition was one of the most prominent applications driving research in time-series analysis
  – Need to match time-series with each other however tradition metrics such as Euclidean distance are brittle to time-series noise and common variances (missed measurements)

• Time-series data tends to introduce both measurement noise and variance (possibly worse under IoT devices due to cheaper sensors)
  – Irregular sampling rates (measurements do not align)
  – Noisy measurements (missing measurements, false or perturbed measurements)

• Tasks such as recognition needs to be able to handle these issues
Challenges in Gait Recognition

Time-series signals generated by gait is periodic, steps generate the same pattern
For recognition, one could use long or short sequences
However, there are challenges to overcome:
• Unlike gesture recognition, the start and end point is less clearly defined
  – Time-series alignment
• Variations in pace cause the signal to be compressed or stretched
• Variations between walks and days (sensor placement, terrain, shoes, mood)
Approach

Could use template matching

- Requires strong alignment constraints (Gait Period Detection)
- Need to be careful selecting length of segment for comparison

- Need flexible similarity metrics to manage variations present (alignment/shift, stretch)
- Possible metrics:

Convolution – $\max$ (sliding Euclidean)  Dynamic Time Warping (DTW)
• Obviously, we would like to diminish segment length constraints
• Solution: use distance metrics and encode data in proper representations, called features, suited for recognition
• Two flexible similarity measures for time-series:
  – **Convolution** - reduces the necessity of finding every step within gait signal (determines the optimal alignment of segments)
  – **DTW** - provides similarity under stretching and compression and partial alignment
Convolution

- A discrete convolution between two signals, $f$ and $g$ as:
  \[(f \ast g)[n] \overset{\text{def}}{=} \sum_{m=-\infty}^{\infty} f^*[m] \ g[m + n].\]

- Take the inner product at every offset position for signals $f$ and $g$. 
Dynamic Time Warping

- Used to measure similarity between time-series data where measurements vary in time and speed (i.e. stretching/compression)
- Goal is to find optimal non-linear mapping (warping function) in time between two signals, $f$ and $g$
DTW cont.

As we have seen, we need to find the optimal transformation that provides the minimal cost

\[ P = c_1, c_2, c_3, \ldots, c_k \quad \text{where} \quad c_k = (f_i, g_j) \]

\( P \) is a warping path corresponding to alignments of points of \( f \) and \( g \) (with constraints, monotonicity, continuity, boundaries)

\[
DTW(f, g) = \min_{P} \sum_{K} cost(c_k)
\]

\[
\text{cost}(c_k) = \text{cost}(f_i, g_j) = ||f_i - g_j||
\]
DTW cont.

DTW can be computed using dynamic programming using the following recurrence equation:

\[ h(1, 1) = cost(f_1, g_1) \]

\[ h(i, j) = \min \begin{bmatrix}
  h(f_i, g_{j-1}) + cost(f_i, g_j) \\
  h(f_{i-1}, g_j) + cost(f_i, g_j) \\
  h(f_{i-1}, g_{j-1}) + cost(f_i, g_j)
\end{bmatrix} \]

- \( h \) corresponds to filling in an \( N \times N \) table
- DTW similarity is the minimal path from \( h(n,n) \) -> \( h(1,1) \)
DTW Example for Gait Signals

Dynamic Time Warp of Subject 5

- Blue: Normal Pace
- Red: Fast Pace

(time)
Feature Encoding via Flexible Similarity Metrics

• Construct a dictionary, $D$, each column is time-series measurements of a given length

• We define a feature encoding as: $f_i = \text{dist}(D_i, x)$

• $D_i$ is the $i$-th column of $D$
  – Dist() can be any suitable distance function for two vectors

• Geometrically, each column can be though of as anchor points within space, each providing a relative distance measure for each segment
  – The new feature representation then can be used with a linear classifier (SVM, logistic regression, decision tree, etc.)
Experiments

- Currently 9 subjects
- Types of walking sessions
  - Walks the same day, between days, different paths, different paces
- Accelerometer, gyroscope on Android smartphone in front left pocket
  - Also have accelerometer readings from smart watch on left wrist
- Dataset is available for use (and always looking for new subjects to record!)

<table>
<thead>
<tr>
<th>setting</th>
<th>$\text{Enc}_{\text{conv}}$</th>
<th>$\text{Enc}_{\text{conv}} + \text{aug}$</th>
<th>$\text{Enc}_{\text{dtw}}$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>complete training set</td>
<td>.99</td>
<td>.99</td>
<td>.99</td>
<td>.94</td>
<td>.90</td>
</tr>
<tr>
<td>different sessions</td>
<td>.96</td>
<td>.95</td>
<td>.65</td>
<td>.90</td>
<td>.88</td>
</tr>
<tr>
<td>different orientation</td>
<td>.88</td>
<td>.92</td>
<td>.73</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>different days</td>
<td>.69</td>
<td>.77</td>
<td>.76</td>
<td>.48</td>
<td>.53</td>
</tr>
<tr>
<td>different pace</td>
<td>.48</td>
<td>.70</td>
<td>.79</td>
<td>.75</td>
<td>.76</td>
</tr>
</tbody>
</table>
Towards IoT Security Solutions

• Gait recognition is promising for user identification purposes
• For low-risk applications, gait or gesture base identification provides a feasible solution for IoT
  – Difficult to mimic biometric data (gait, voice, etc.)
• For higher security constraints, several mechanisms could be layered
  – Leverage gait, voice, and environment sensing tools together for improved security
Nymi - Biometric Authentication Device

• Device acts as keystore for the user
  o Allows for authentication with other devices and secure transactions

• Authentication process
  o User presses electrode which starts capturing the ECG signal
  o Nymi sends signal over secure channel to Authorized Authentication Device (AAD) - the user’s phone/computer
  o AAD matches the new sample against known good template

• User remains authenticated until device is removed (ECG only used during authentication process)

• Also has accelerometer/gyroscope and proximity sensor
“Analysis of Human Electrocardiogram for Biometric Recognition”

- Preprocessing
  - Bandpass filter used to remove low/high frequency noise and corrects baseline wander

![Graph showing baseline wander](image-url)
“Analysis of Human Electrocardiogram for Biometric Recognition”

• Feature Extraction - Two Approaches
  • Use analytic features based on “fiducial detection” - old method
  • Use Autocorrelation to find repeating patterns in the signal and DCT for dimensionality reduction - their approach
“Analysis of Human Electrocardiogram for Biometric Recognition”

---

**Figure 2:** Block diagram of proposed systems.

**Figure 6:** (a–b) 5 seconds window of ECG from two subjects of the PTB dataset, subject A and B. (c–d) The normalized autocorrelation sequence of A and B. (e–f) Zoom in to 300 AC coefficients from the maximum for different windows of subject A and B. (g–h) DCT of the 300 AC coefficients from all ECG windows of subject A and B, including the windows on top. Notice that the same subject has similar AC and DCT shape.
“Analysis of Human Electrocardiogram for Biometric Recognition”

• Classification
  o Uses K-means to find centroid (Euclidean Distance)
  o New samples compute their distance to each centroid and are assigned the label of the closest one
Nymi authentication

• Continuously sends ECG signal to AAD until authenticated or timeout occurs
• Thoughts on false positive (FAR) versus false negative (FRR) in this case?

Figure 1. Percentage true accept rate vs. authentication time.
Communication between Nymi and AAD

• Uses Bluetooth low-energy radio for communication (used in many smart devices like fitbit, door locks, etc.)
• Nymi white paper cites “Bluetooth: with low energy comes low security”
  o Recall this paper outlines both how to do passive sniffing and injection attack
• Nymi protects against this by having hardware capable of ECDSA (Elliptic Curve Digital Signature Algorithm). Uses Diffie Hellman key exchange when it wishes to communicate with a new AAD or when sending information to a 3rd party to provide user authentication. All traffic sent from/to the device is encrypted. Private keys are stored on the device