Security and Privacy in DR: Sensors and Behavior Inference

Umesh Shankar

UC Berkeley

Two Main Problems

1. Sensing/Actuation at customer site
   - Wireless sensors/actuators
   - Detailed data may be stolen
   - Actuators may be misused
2. Data Sent from customer to utility
   - Privacy compromise
   - Customer activity may be inferred
   - **Main Future Work**

Part I: Sensors

Why sensors?

- Take readings from more places to get finer control
- Also better/quicker response
- Deploy cheaply and easily

Insecurity Inheritance

- Sensor nets inherit problems from:
  - Ordinary networks
  - Wireless networks
  - Ad-hoc networks
- Plus add their own challenges due to resource constraints

Attacks

- Physical
- Network layer
- Routing layer
- Application

Physical Attacks

- Power draining
  - Communication is costly
  - DoS attacks can kill nodes
- Physical node capture / cloning
  - Node replication is feasible
  - (Improved hardware can change that)

Network Layer

- Insertion / deletion / modification of packets
- Insertion: use crypto authentication
- Deletion/Modification: require jamming
  - Jamming is “dumb” but takes power

Routing / Route Formation

- Threat model
  - Can use out-of-band channels, more powerful nodes (e.g. laptop)
  - Can capture, clone legitimate nodes
- Attack classes
  - Induce bogus routing trees to be formed
  - At runtime, selectively forward/drop/modify packets

Some Attacks

- Sinkhole (attract traffic to gain control)
- Sybil (multiple identities)
- Wormhole (out-of-band routing)
  - Simulates node cloning
- HELLO (advertise good routes)
- ACK spoofing
- Rushing attack

App Layer (Data Collection)

- Compromised node readings
- Simply faulty nodes
- Certain functions can’t be computed accurately, e.g. mean, min, max
- Adversary can arbitrarily influence computed value of these functions

Broad Recommendation

- Keep it simple!
  - Complexity is the enemy of security
- Eliminate unnecessary features
- Use standard, well-understood techniques

Recommendations (I)

- If physical capture is a threat:
  - Tamper-resistant HW, Plug-in power
- Radio: Spread-spectrum (DSSS / FH)
  - Protects against incidental interference
- Routing:
  - Don’t route! (Single-hop network)
  - If not single-hop, then, fixed routing tree

Recommendations (II)

- Crypto use:
  - Randomness (semantic security)
  - Timestamping (replays)
  - MAC (non-malleability)
  - **Use standard codes and protocols!**
- Data processing
  - Use resilient aggregates
  - Median, trimmed average, etc.

Part II: Privacy

What is the problem

- Power usage patterns can reveal customer behavior
  - Both legal (sleep/wake patterns) …
  - …and illegal (marijuana growing)
- DR requires more fine-grained meter reading, revealing patterns
- Data theft at utility is now much more damaging

The Best Solution for Privacy

- Intelligent Endpoints
- The meter computes the bill
- Doesn’t send back detailed usage

$\Rightarrow$ **Unavailable data can’t be misused**

What if some details needed?

- More detailed information may be used to learn usage patterns and issue better guidelines
- This can be handled using voluntary subjects and controlled sampling
  - Like Nielsen ratings

Another in-between option

- Use *aggregation* and *anonymization* of data
- Removes identifying information, but still yields useful data

Examples

- Anonymizing across a block yields useful info, but preserves some anonymity
- Aggregating a block gives some estimate of a house on it, esp. combined with the amout of its bill (assuming a normal distribution)

Anonymization: Mixing

- Used in anonymized routing and electronic voting
- Remove identifying information
- Send into the mix; it gets encrypted and randomly routed around
- We have results on resulting anonymity guarantees

Measuring anonymity

- Anonymity uses an entropy measure
  - Compute probability distributions for estimated customer usage
  - Entropy is a function of distribution
  - Roughly: how precise is our guess?

Two Axes of Aggregation

- Aggregate across multiple customers
  - Sum up a neighborhood's usage
- Aggregate over time
  - Only show averages over a day
  - Already performed at small scales (minutes or hours)
- These can be combined

Anonymity from aggregation

- If we know the distribution of usage curves and local averages, how much can we guess about a customer?
- How does geographic aggregation compare with temporal aggregation?

Future Work

- Quantify the privacy guarantees of various anonymization/aggregation techniques
- Look at techniques in secure databases

Thanks!

Contact:
ushankar@cs.berkeley.edu