Critical Infrastructures
(of today and tomorrow)

Simin Nadjm-Tehrani

www.ida.liu.se/~rtslab

Department of Computer & Information Science
Linköping University, Sweden

and

University of Luxembourg
Outline

• What are the challenges to today’s critical infrastructures?

• Overview of the emerging infrastructures
  – A post-disaster communication network as an extreme case

• Own work in 2003-2005 in one of the first European projects on critical infrastructures
Attributes of dependability

Availability
- Readiness for use

Reliability
- Continuous correct service

Integrity
- No unauthorised change
Challenge 1

Complexity and interdependencies
• Classic regional power grid
• One vertically integrated utility
• Grid control by frequency following
• Dedicated communication
Restructuring of the Grid

• Deregulation: organisations can enter into bilateral or multilateral power generation contracts
  – Large scale operation: from centralised to distributed control
  – Difficulty of coordination among independent service operators
• Approaching grid capacity
• New monitoring and control problems
Local impact spreads

Provider A

Provider B

Provider C

Load following?

Distribution Substations

Customer X

Provider B

Provider A
Need for communication & trust

• Line frequency can no longer be the implicit communication channel
• Ideally contracts and capacities need to be known to everyone for cooperative control

In reality ...
• No operator wants to disclose information unless mandated by authorities
• Line frequency not enough for stabilisation: one needs to know the state of equipment, detailed load profiles, pricing,…
• 646 flights delayed as a direct result of a failure in a communication link that transmits flight plan data from the Georgia facility to a similar facility in Salt Lake City

• Flights from a wide swath of the United States, from Dallas and to the East Coast delayed

• The FAA: the source of the computer software malfunction was a "packet switch" that "failed due to a database mismatch."
Transition from managed to unmanaged
• Skype today provided a few more information pieces about the reasons behind its massive network outage last week.

• The network outage was initially caused by a "massive restart of [its] user's computers across the globe within a very short timeframe as they rebooted after receiving a routine software update."

• That high number of reboots was followed by an equally high number of log-in requests, which resulted in what Skype calls a "chain reaction."
Challenge 3

Heterogeneity
Converging networks

- From cellular …
- … to adhoc networks with no infrastructure
- … to multi-region intermittently-connected networks
Reliance on novel technologies

• Wireless Communication
  – Almost taken for granted as part of the infrastructure today
  – GPRS, HSDPA, WiMAX, Wi-Fi, …

• Distributed cell networks
  – Local (per customer) generation of electricity
  – Dynamic energy market trading at customer level
Organised threats with economic motives or adversary disruptions
Symantec Threat Report - Dec 07

- An average of 61,940 active bot-infected computers per day in the second half of 2007, an increase of 17% from the previous period.

- 499,811 new malicious code threats were reported to Symantec, a 136% increase over the first half of 2007.
Georgian govt. web attacks

• Gadi Evron, a prominent Internet security researcher and the founder of Israel's Computer Emergency Response Team, posited that the attackers are more likely nationalistic "enthusiasts" than organized criminals or Russian government operatives.
<table>
<thead>
<tr>
<th>Challenge</th>
<th>Emerging solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complexity and interdependencies</td>
<td>Modelling, Risk analysis, Provisioning</td>
</tr>
<tr>
<td>Transition from managed to unmanaged</td>
<td>P2P technologies, self-managing systems</td>
</tr>
<tr>
<td>Heterogeneity</td>
<td>Standardised protocols, Overlay networks, Software defined radio</td>
</tr>
<tr>
<td>Organised threat, fraud and disruptions</td>
<td>Hardening, Intrusion tolerance, diversity, partial rejuvenation</td>
</tr>
</tbody>
</table>
Outline

• What are the challenges to today’s critical infrastructures?

• Overview of the emerging infrastructures
  – A post-disaster communication network as an extreme case

• Own work in 2003-2005 in one of the first European projects on critical infrastructures
What happens in the worst case?

- Existing infrastructure collapses

Chaotic & surprising
Network: lack of resources

Time is running out...

- Actors are spread out and mobile
- Communication culture clashes
Our hypothesis

• Hastily formed networks can have a role to play

• Use commodity hardware and massively distributed software

• Have built-in mechanisms for
  – When batteries are in short supply
  – Mobility changes connectivity
  – Dealing with overload and urgency
  – Detect and respond to abuse
Multiple information owners/users
## Hastily formed networks

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Emerging solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disconnectivity as a norm</td>
<td>Store-and-forward techniques, delay-tolerant networks (DTN)</td>
</tr>
<tr>
<td>Resource constraints</td>
<td>QoS optimisation techniques, prioritisation</td>
</tr>
<tr>
<td>Infeasibility to centrally manage</td>
<td>Gossip-style distributed protocols</td>
</tr>
<tr>
<td>Heterogeneity</td>
<td>Overlay networks, DTN bundles</td>
</tr>
<tr>
<td>Less organised opportunistic threats, adversary disruptions</td>
<td>Reputation-based systems, Selfish-resistance protocols, Decentralisation</td>
</tr>
</tbody>
</table>
Outline

• What are the challenges to today’s critical infrastructures?

• Overview of the emerging infrastructures
  – A post-disaster communication network as an extreme case

• Own work in 2003-2005 in one of the first European projects on critical infrastructures
• Goal: to enhance survivability of Large Complex Critical Infrastructures (LCCIs)

• Electricity and telecommunications networks as practical examples

• Granted pre 9/11!

• Ended in 2004

plus a panel of senior government and industry advisors
Challenges

General:
• Increase information quality for administrator
• Recognise unknown attacks
• Predict future overloads

Telecom specific:
• Decrease no. of alarms
• Decrease false positives (higher availability)
The Safeguard approach

Safeguard agents
Safeguard architecture

Other LCCIs

Negotiation agent

HMI Agent

Topology agent

Correlation agent

Action agent

Diagnosis wrapper

IDS wrapper

Anomaly detector agent

Actuator

Home LCCI

High-level agents

Low-level agents
Anomaly Detection

- ADWICE: Anomaly Detection With fast Incremental Clustering

- Joint work with Kalle Burbeck

- Not a silver bullet: part of the larger Safeguard context
Intrusion detection

- **Protected System**
- **Attacker**
- **Intrusion Detection System**

**Misuse Detection**
- Normal Behaviour
- Model

**Anomaly Detection**
- Normal Behaviour
- Model
- Bad Behaviour
- ADWICE uses clusters to represent normality
- Adaptation of an existing data mining algorithm (BIRCH)
What is a data point?

- General: A set of numeric values
  - E.g. measurements from sensors

- What about IP packets?
  - A vector of alphanumeric values in header of an IP packet
  - Transformed into vector of numeric values
  - In our tests: 41 dimensions

- Need efficient storage and search among summaries of collections of data points
Basic ADWICE concepts

- **CF (Cluster Feature)**
  - Summary of cluster
  - [No, Sum, Sum of sq]

- Index: CF Tree

- Maximal number of clusters (M)

- Threshold requirement (TR)

- Branching factor (B)
Efficient operations

• We have: $\text{CF} = \left\langle n, \sum v_i, \sum v_i^2 \right\rangle$

• Can compute the Centroid $v_0$:

$$\sum v_i / n$$

• Can compute the Radius:

$$\sqrt{\sum (v_0 - v_i)^2 / n}$$
Threshold:  
Max Number of Clusters: 3
Branching factor: 2

Data Space  
CF Tree
Threshold: 0

Max Number of Clusters: 3
Branching factor: 2

Data Space

CF Tree
Threshold: 3
Max Number of Clusters: 3
Branching factor: 2

Data Space

CF Tree
Threshold: 3
Max Number of Clusters: 3
Branching factor: 2
ADWICE detection

Threshold: Max Number of Clusters: 3
Branching factor: 2

Data Space

CF Tree
Evaluation

- KDD99 Data
- General properties
  - Session records (TCP/UDP summaries)
  - 41 features (flags, service, traffic stats ...)
- Training data
  - 4,898,431 session records
  - 972,781 normal, the rest (attacks) not used
- Testing data
  - 311,029 session records
  - normal data and 37 different attack types
Detection rate vs. false positives

Detection rate

False positives rate

ADWICE-DT-TRR

ADWICE-DT-TRD
Index errors

• Some false positives are due to index errors
ADWICE-Grid

• A new version of the algorithm: separates cluster formation and index updates

• How does ADWICE-Grid work?
Threshold: [0.0, 0.2] [0.8, 1.0]

Max clusters in Leaf: 2
ADWICE-Grid: Training

Threshold: (Search width)

Max clusters in Leaf: 2

Data Space

CF Tree

Max clusters in Leaf: 2
ADWICE-Grid: Detection (1)

Data Space

CF Tree

[0.0,0.2] [0.6,0.8]
ADWICE-Grid: Detection (2)
Detection rate vs. false positives

Source: [Burbeck & Nadjm-Tehrani 04,07]
• Anomaly detection may produce many similar alarms (e.g. DoS, Probes, False positives)
• Similar alarms can be aggregated without losing accuracy

Alarm aggregation

Normal alarms: \(<t_1, \text{HTTP}, \ldots> \quad <t_2, \text{HTTP}, \ldots>\)

Aggregated alarm: \(<\text{Start, End, Count = 2, HTTP}, \ldots>\)
Alarm aggregation results

![Graph showing alarm aggregation results](image-url)
Safeguard 100+ test network

Pseudo-External Network

Internet

Safeguard Systems

Workstations

Server Zone

Safeguard

c7000

c2500

stns6

stns77

stns78

Safeguard workshop
Estonia
August 28, 2008
One HMI agent interface
A Safeguard scenario

Malicious
User
Scripts

Number of alarms

Period number (1 minute per period)
Correlating alarms

Multi sensor correlation

Adaptive Filtering

Aggregation

Wrapper

Normalisation

Static filtering

Samhain

Syslog

Model Adaptation

Anomaly Detection

Whitelist Filtering

TcpDump
Need for normality adaptation

- Normality is not static!
New cases of normality

- Normality changes
  - New type of normal behaviour
- Old model incomplete
  - Evaluation using KDD data gives ~300 false positives for new normality
Evaluation of normality adaptation

- Admin or system reacts
  - Recognize new false positives
  - Tells ADWICE to learn this behaviour
- Normality model adapted
  - From 300 to 3 false positives!
Forgetting

- System keeps track of model usage
  - If time since last usage is very long for subset of clusters
  - Decrease size (influence) of those clusters and finally remove them if not used
Safeguarding critical infrastructures needs:

• Adaptive elements
• Incremental and scalable algorithms
• High performance for large volume of data

• Demonstration on realistic test beds
  – Research on open data sets :-) 

• Understanding and mitigating interdependencies
Current track

- Application of ADWICE in anomaly detection for water management systems
  - Cooperation with Environment Protection Agency (EPA), USA
  - Time series data from simulated water system over an interval of one week

- Talk to me if interested to join!