THE UNIVERSITY OF MELBOURNE

Detecting Attacks on Networked Services

Christopher Leckie
Academic Centre of Cyber Security Excellence
School of Computing and Information
The University of Melbourne

Oceania Cyber Security Centre
caleckie@unimelb.edu.au

OCSC
Oceania Cyber Security Centre

Overview of ACCSE

Multi-disciplinary centre for education, training, outreach, research

Department of Education and Training

Commenced
July 2017

Academic Centre for Cyber Security Excellence

Computing and Information Systems
Electrical and Electronic Engineering
Arts, Law

Affiliated with
Melbourne Networked Society Institute
Research partnership with NBNCo

Overview of OCSC

Engage with industry in research and training on cyber security

Victorian State Government

Victorian Universities
(Monash, U.Melbourne,
Deakin, RMIT, Swinburne,
Federation, Latrobe, Victoria)

Commenced
October 2016

Co-located with Data61
International Collaborations: Oxford, Tel Aviv

ACCSE: Aims and Objectives

- **Identify new opportunities for engagement with industry & govt on research and training, particularly in partnership with the Oceania Cyber Security Centre (OCSC)**
- **Expand opportunities for cyber security focused PhD and Masters students to work with industry through internships**
- **Tailor research training resources to reach wider audience through professional and executive education**
- **Introduce a new Cyber Security specialisation in the Master of IT**

Examples of Research Expertise

- **Security analytics**
- **Detecting attacks in large, complex systems**
- **Business information systems security (risk assessment and incident response strategies)**
- **Formal methods for security and verification to design resilient platforms**
- **Privacy-preserving data sharing among organisations**
- **Psychology of cyber crime**

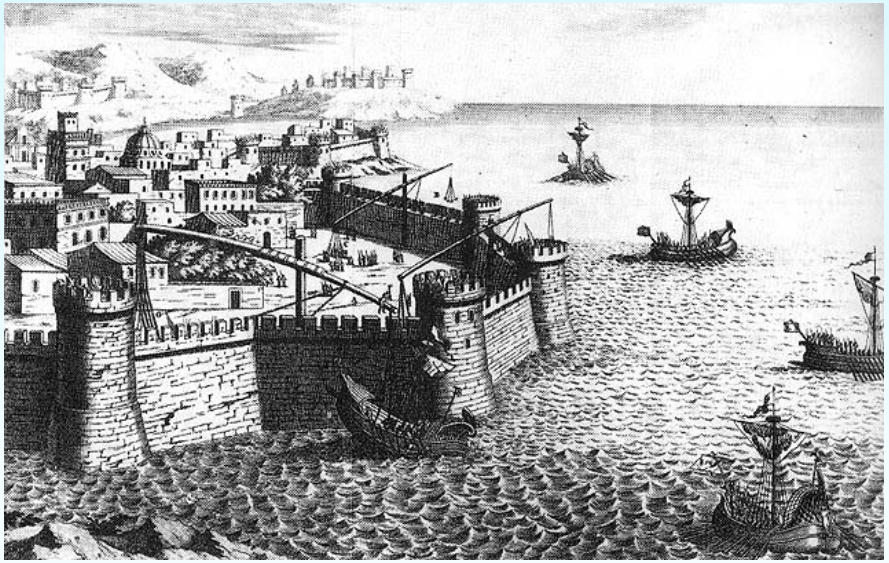
Context


BBC
News
Sport
Weather
Shop
More
NEWS
Sections
Technology
Cyber attacks briefly knock out top sites
21 October 2016
Technology


NBC NEWS
SECTIONS
NIGHTLY NEWS
MEET THE PRESS
DATELINE
NEWS > U.S. NEWS
WORLD
INVESTIGATIONS
CRIME & COURTS
LATINO
NBCBLK
NEWS OCT 22 2016, 9:43 PM ET
DDoS Attacks That Caused Chaos on Web Were 'Sophisticated': Dyn
by JAVIER E. DAVID, CNBC

Denial of service attack

Prevent legitimate users from accessing a critical resource

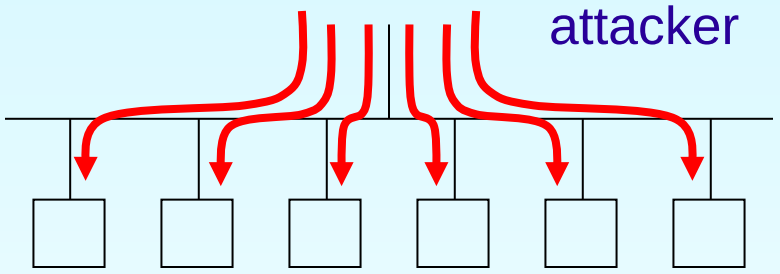


https://www.cs.drexel.edu/~crorres/bbc_archive/secrets.html

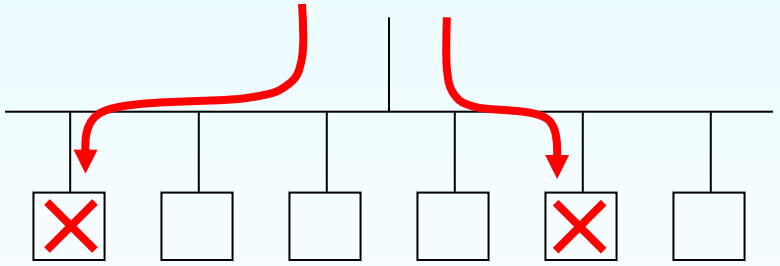
http://www.vintersections.com/2011/02/philippines-celebrates-25th-anniversary.html

Typical Network Attack Scenario

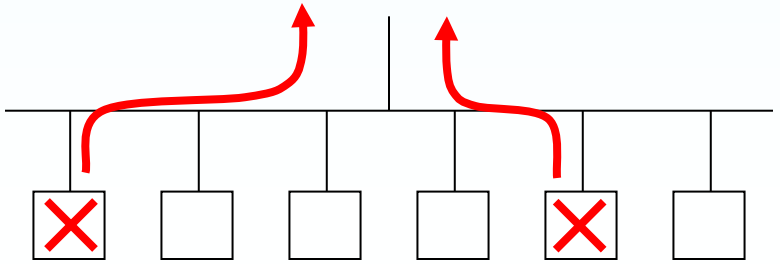
1. Scan for hosts with known weaknesses



2. Gain access to hosts and install attack software

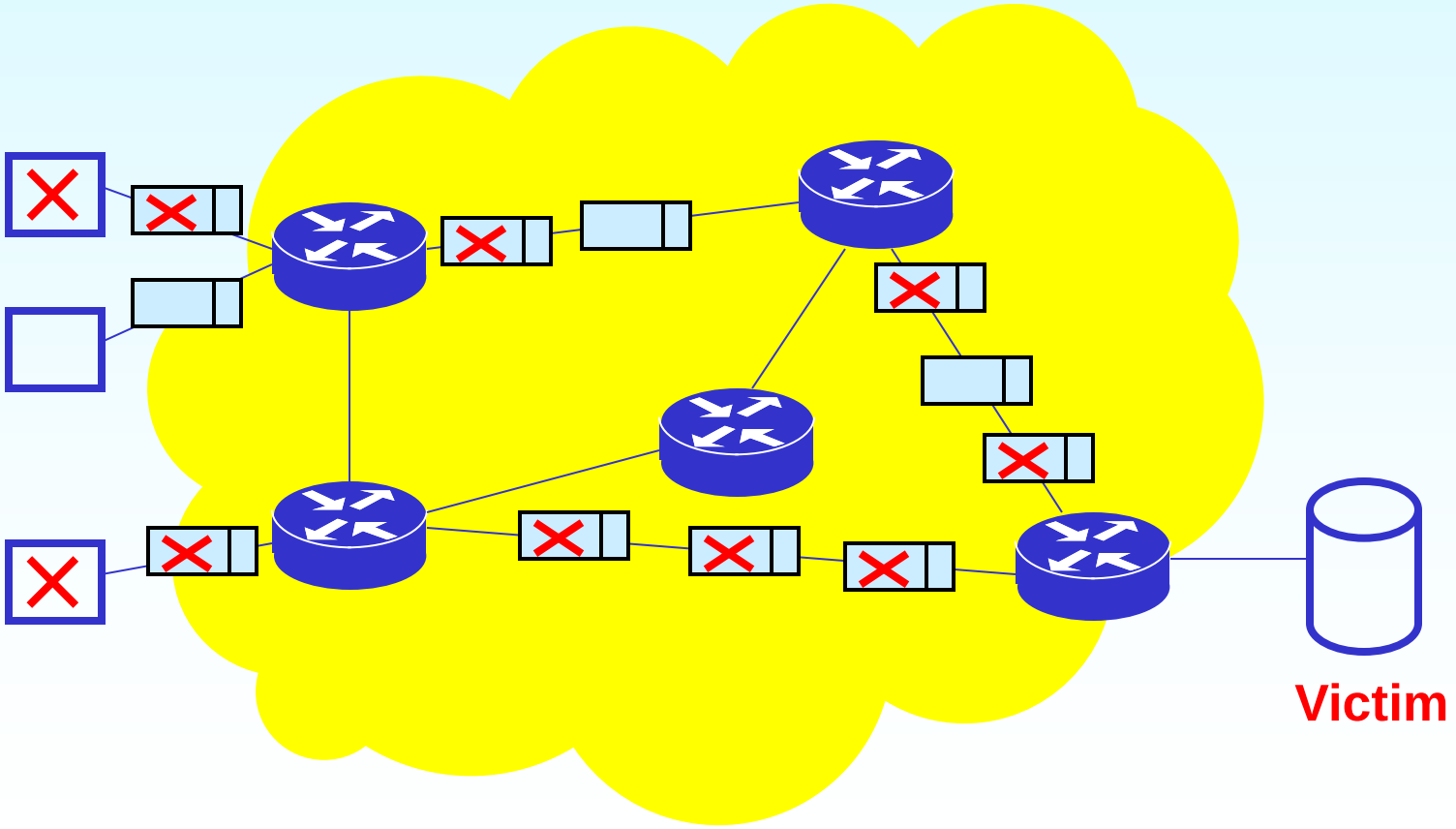


3. Use infected hosts to launch denial-of-service attack

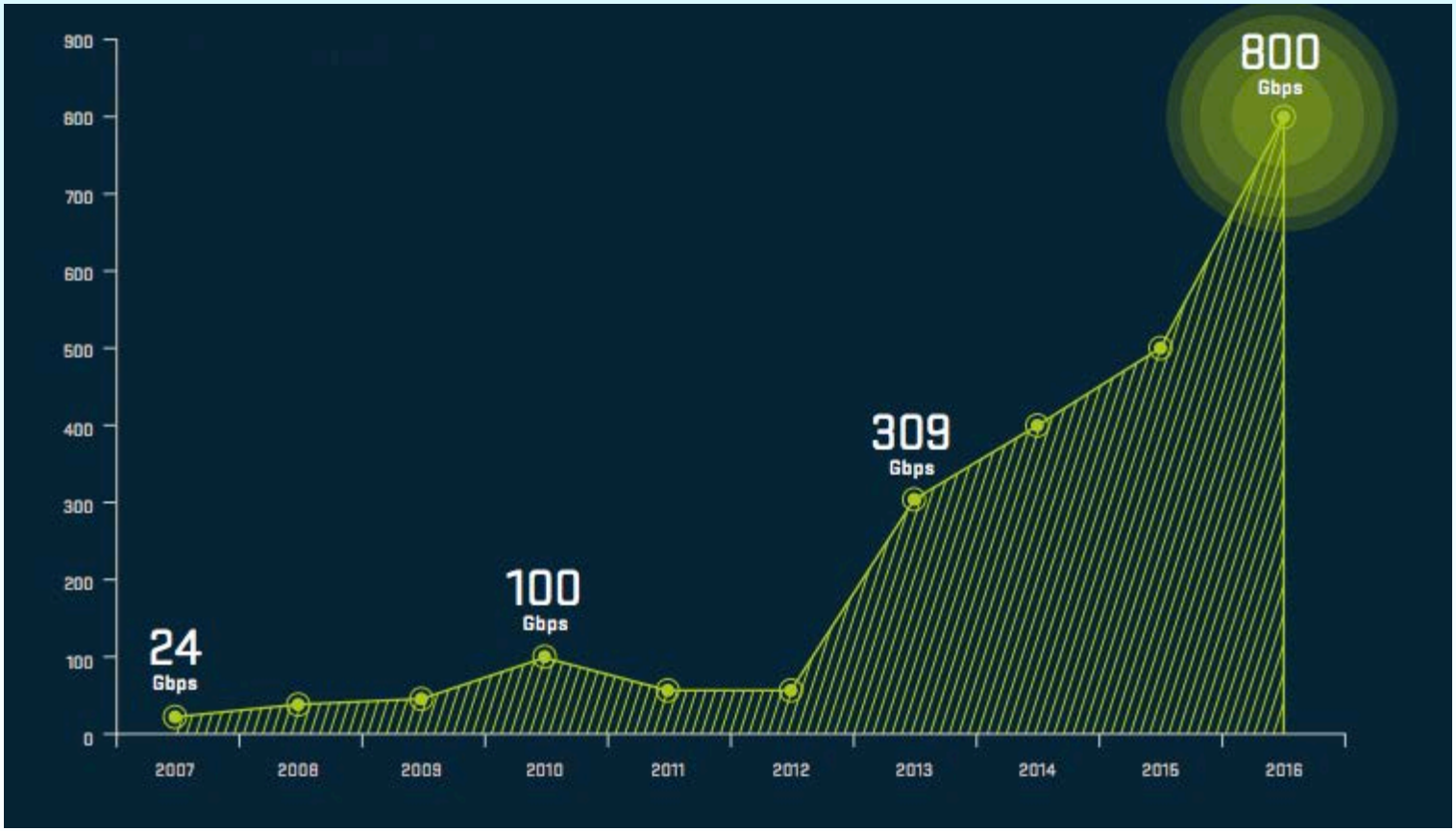


Distributed Denial of Service Attacks

Exhaust resources of victims using malicious traffic



The Growing Arms Race



Trend in maximum DDoS attack rate over past 10 years

[Source: Arbor 12th Annual World Infrastructure Security Report, 2017]

Impact of DDoS Attacks – Examples from Finance Sector

Operation Ababil (commenced Sept 2012)

- **DDoS HTTP and DNS attacks from up to 20,000 sources**
- **Multiple financial targets (NYSE, J.P. Morgan Chase)**
- **Politically motivated**
- **Total of 249 hours downtime of major U.S. bank websites over 6 week period in 2013 [Source: Keynote Systems]**

More recently: HSBC online access (Jan 2016)

What is the risk to the utilities sector?

General Approaches to Defence

- **Absorption**
  - Bigger infrastructure; Content Delivery Networks
- **Filtering or scrubbing traffic**
  - On-premises or using cloud-based service
- **Classification of traffic source**
  - e.g. application traffic from non-human source

Example of Attack on a Utility – Ukraine Power Grid 23 December 2015

Remote attackers shut down part of the Ukranian electricity grid, blacking out 230,000 customers

- Introduced via phishing email to staff containing an infected Microsoft Word document
- Having gained initial access, they spent months accessing and mapping the SCADA network
- At 1530 on 23/12/2015, attackers launched a carefully synchronised attack to trip circuit breakers, lock out control staff, and corrupt firmware
- They also launched a Telephony DoS attack on customer call centres

See https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf

https://www.wired.com/2016/03/inside-cunning-unprecedented-hack-ukraines-power-grid/

DDoS on Telephony

- **Attack on phone numbers,**
  rather than IP addresses or applications
- **Flood of malicious inbound calls**

  911 service in U.S., call centres
- **Disruption of physical world**

  Jan 2016: computerised bomb threats against schools
- **Attack on VoIP infrastructure**

  e.g. CPU-based DoS against SIP registration servers

Observations on Ukraine Power Grid Attack

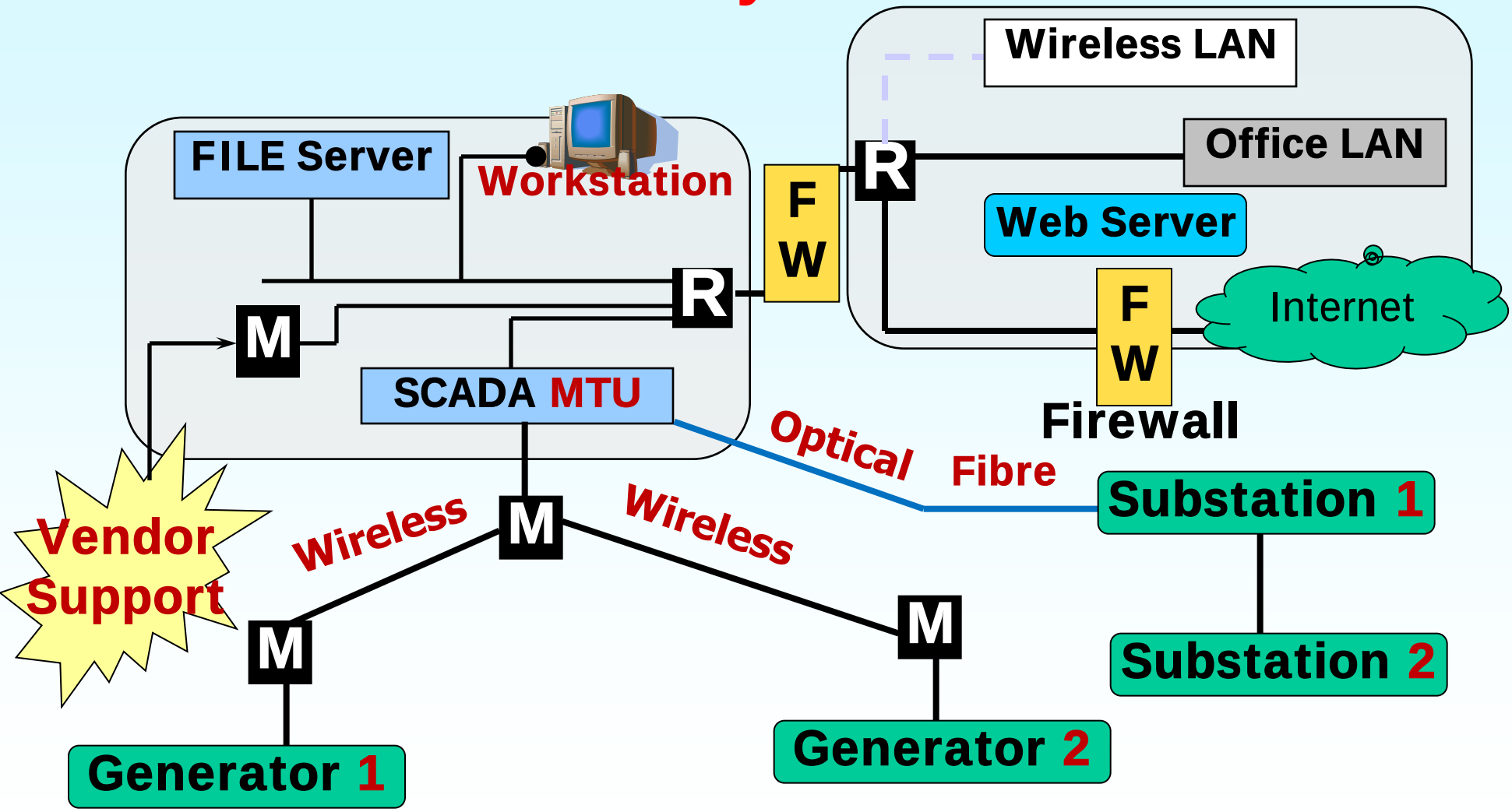
- DDoS was not central to the attack

  → they could trip the circuit breakers directly from the inside

- Possible roles of telephony DDoS?

- If the SCADA network is isolated from the public Internet and corporate network, is there an “attack surface” for DDoS?

Example of SCADA Network for Electricity Distribution



Targets for DDoS in Utilities

Current targets:

- Customer facing corporate web services (billing, faults, etc)
  → affects corporate reputation
  but won't shut down the Industrial Control Systems

Emerging future targets:

- Smart Grid and Advanced Metering Infrastructure
- Industrial Internet-of-Things (IoT)
- Cloud-based Industrial Control Systems

Smart Grid and Advanced Metering Infrastructure

Advanced Metering Infrastructure provides support for:

(1) Meter reading and maintenance

(2) Real-time pricing and demand response

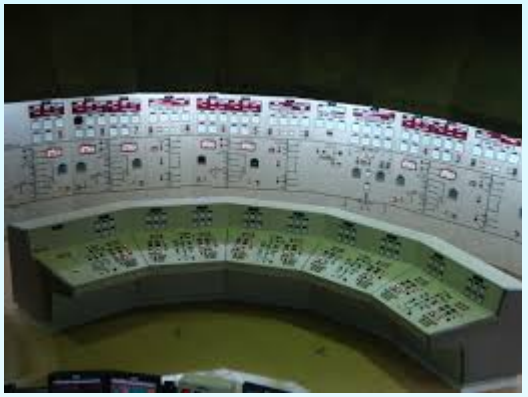
If metering infrastructure can be compromised,
then possible DDoS attack scenarios have been identified
for demand response

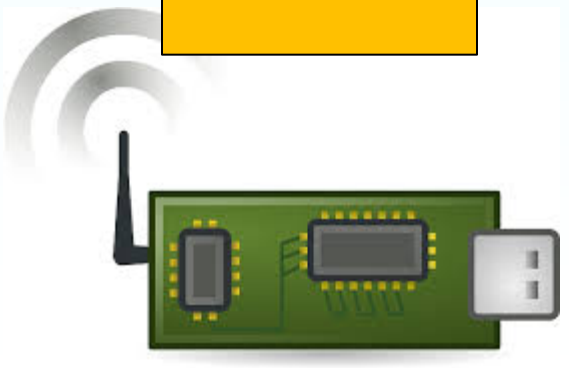
e.g. blocking price signals can disrupt demand response

See NISTIR 7628 “Guidelines for Smart Grid Cyber Security”

EU FP7 SysSec “Advanced Report on Smart Environments”

Industrial IoT for Networked Control Systems





Potential Effects of Attacks in Industrial IoT

- Potential for disruption if wireless (or public) networks are used (even with authentication)
- Filter or fake sensor measurements and alarm messages from customers
- DoS attacks to disrupt communication in delay-sensitive control loops
- Shodan search engine provides ability to search for Internet-connected devices (https://www.shodan.io/)
- Disruption of commercial IoT sensors e.g. thermostats

Adversarial Anomaly Detection

Prameesha S. Weerasinghe
Sarah M. Erfani
Tansu Alpcan
Christopher Leckie
Margreta Kuijper

University of Melbourne
Academic Centre for Cyber Security Excellence

caleckie@unimelb.edu.au

What is Machine learning?

- It is a method of data analysis including making decisions such as classification

How does it work?

- Automatically builds an analytical model by using algorithms that iteratively learn from data
- Machine learning allows computers to find hidden features without being explicitly programmed to extract these features.

Why is it popular now?

- Growing volume and variety of available data
- Increased computational capability
- Affordable data storage

Supervised learning

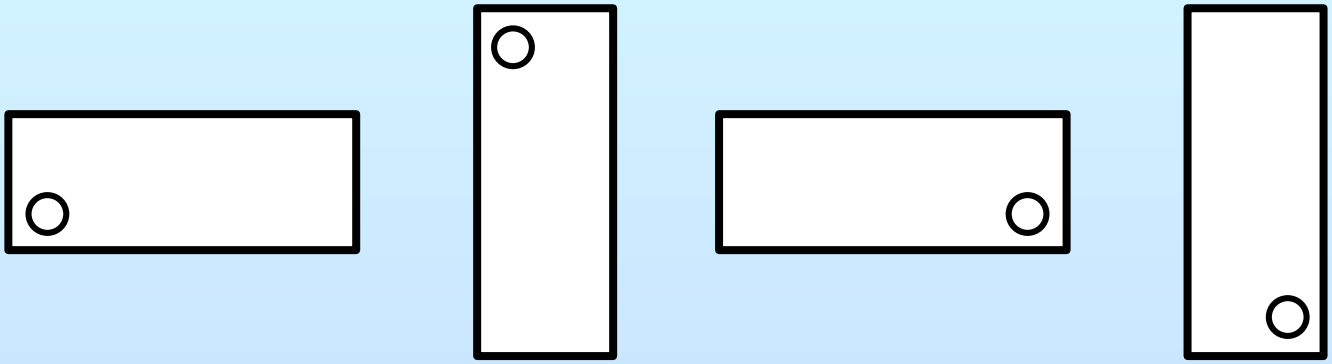
- We give data as well as labels
- The algorithm finds the relationship between the data and the labels - e.g., Classification

Unsupervised learning

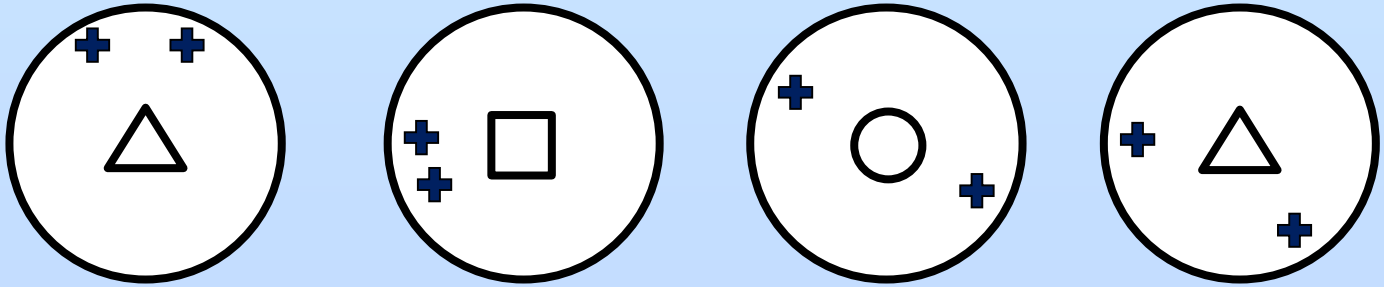
- Data is given without labels
- Algorithm finds patterns in data - e.g., Clustering or Anomaly Detection

Anomaly detection: a general challenge of intelligence?

Spot the odd one out:



a. b. c. d.



a. b. c. d.

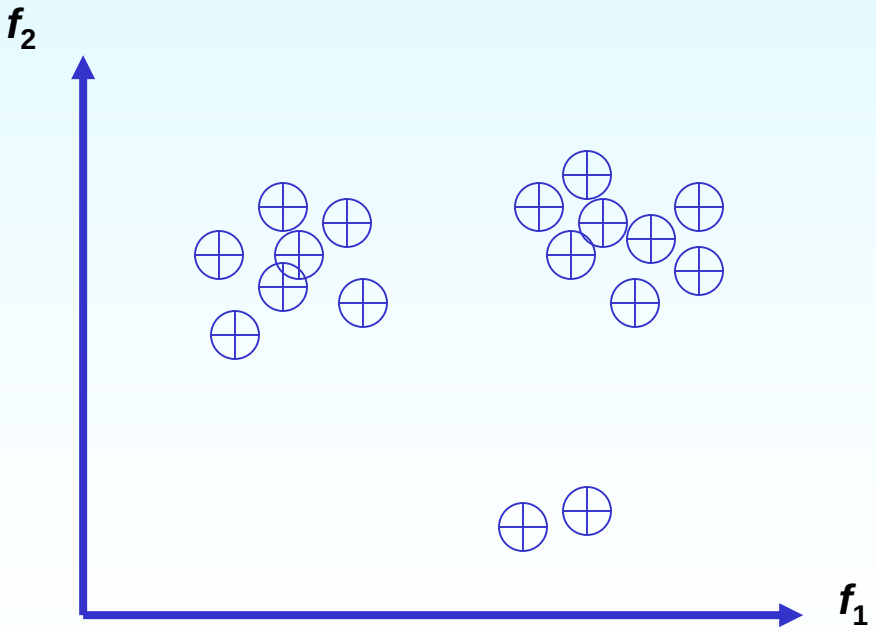
Learning Unusual Patterns (Anomaly Detection)

- Learn a model of “normal” database records
- Use this model to test new records for anomalies
- Any anomalies can be either interesting or errors

Unsupervised Anomaly Detection

[Eskin et al. 2002]

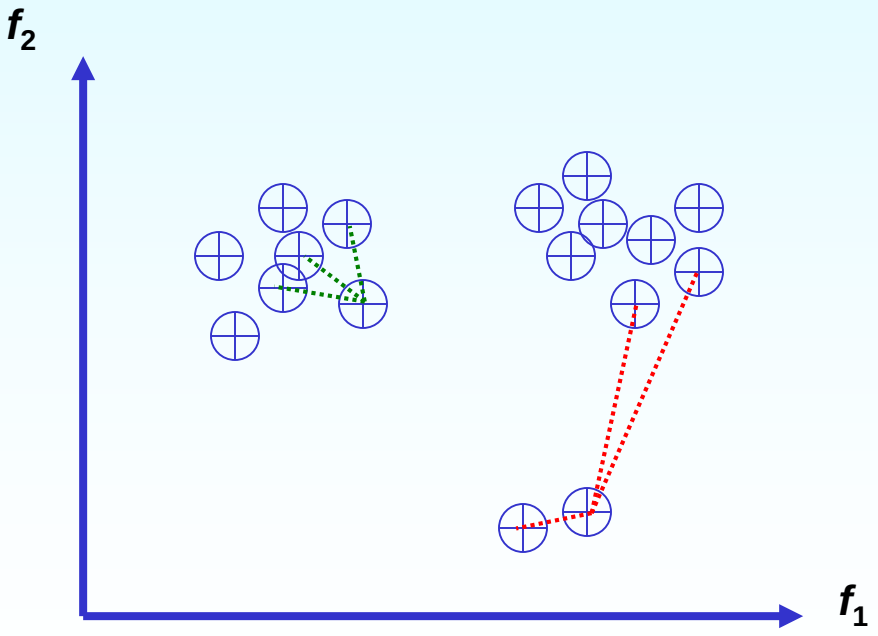
- **Map record fields into a feature space {f_1… f_k}**
- **Cluster similar records**
- **Use large clusters to represent normal records**



Unsupervised Anomaly Detection

K-nearest neighbours:

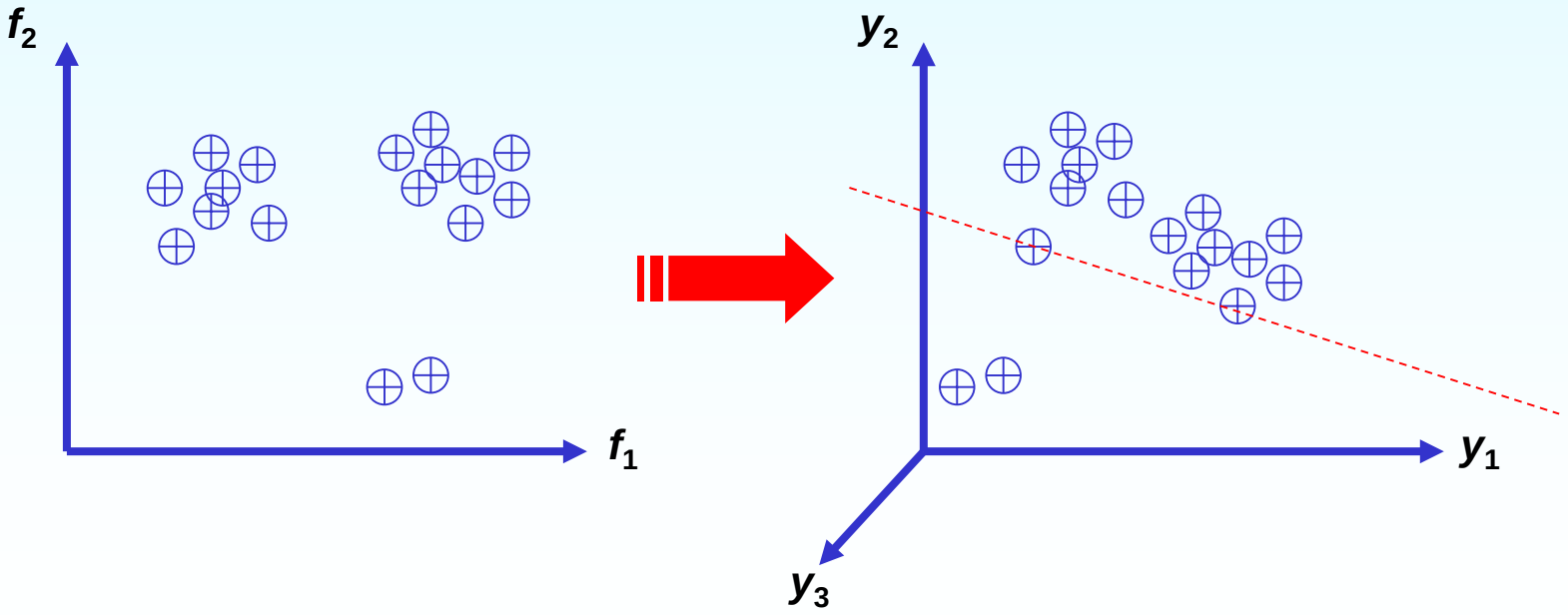
- **Find *k* nearest neighbours of each point**
- **Data points with high kNN distance are in sparse regions of space**



Unsupervised Anomaly Detection

One-class Support Vector Machine:

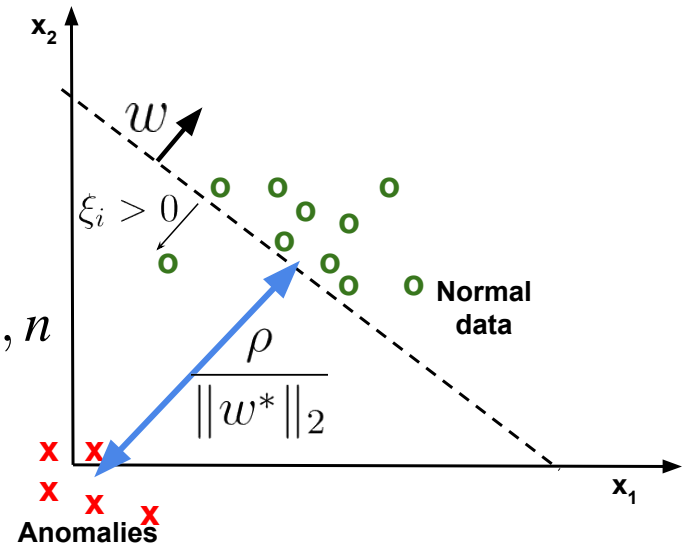
- **Map data points into a higher dimensional space**
- **Find a hyperplane that is *maximally distant* from origin while separating *most points* from origin**



One-Class Support Vector Machines

- An **unsupervised** learning algorithm to **detect anomalies**
- Linearly separates the training data w.r.t. the origin with the highest margin
- The primal optimization problem of OCSVMs is (Schölkopf et al. 2000)

$$\begin{aligned} \min_{w,\ \xi_i,\ \rho} \quad & \frac{1}{2}\|w\|^2 - \rho + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i \\ \text{subject to} \quad & \langle w\,, x_i\rangle \geq \rho - \xi_i,\ \forall i = 1,\ldots,n \\ & \xi_i \geq 0,\ \forall i = 1,\ldots,n \end{aligned} \tag{1}$$



- where $\nu \in (0, 1)$ is the regularization parameter
- take larger value for ν if training set is suspected to be contaminated
- ρ is the offset from the origin
- ξ_i values are the slack variables

Actions of an Adversary

Source: Winnetka Animal Hospital

Can they "poison" our model of what is normal?

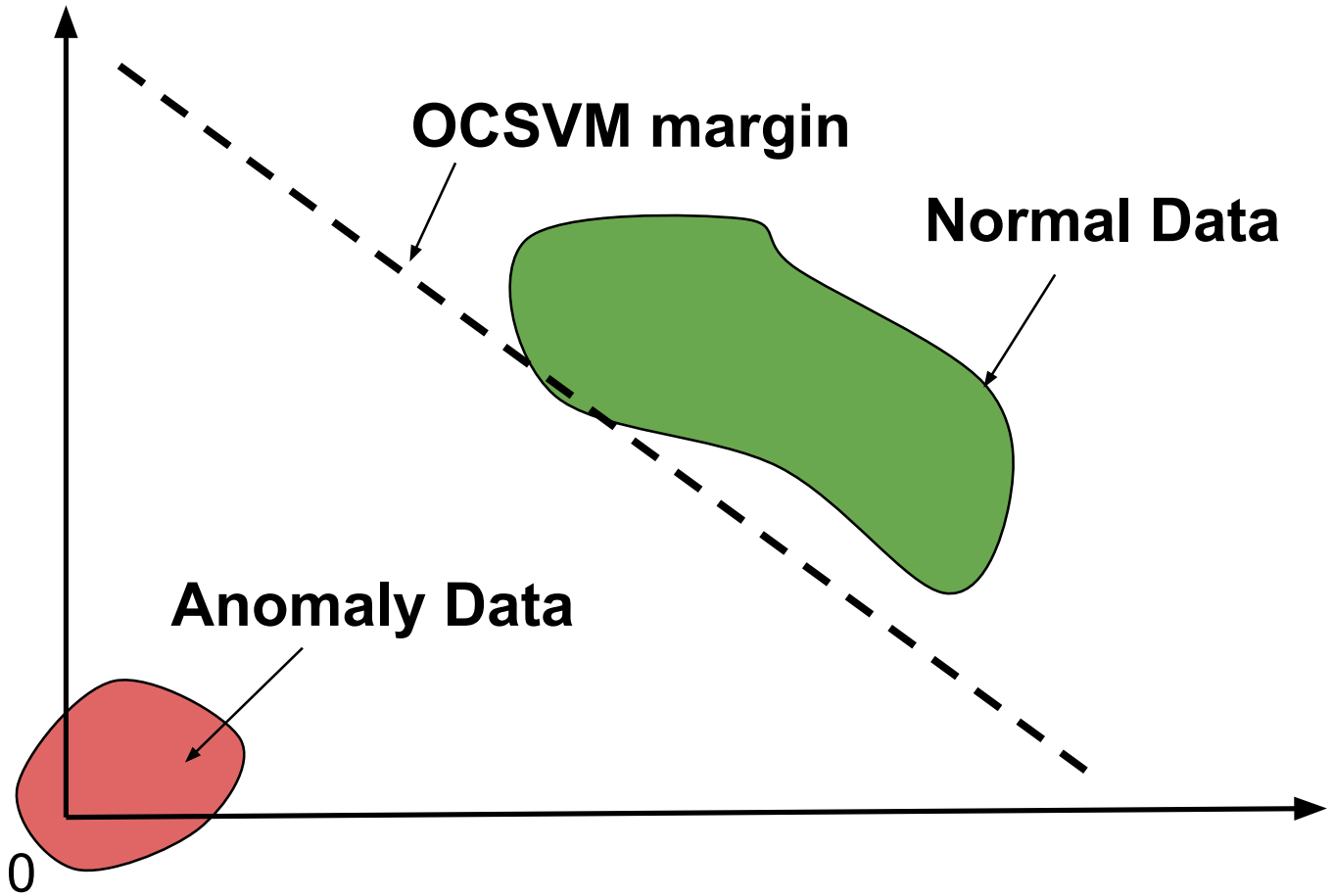
Attack on Integrity

- The ultimate objective of the attacker is to fool the user into labeling anomalies as normal during testing (increase **False Negatives**)
- The attacker would first compromise the classifier by injecting outliers into the training data
- After this, it would be easier for the attacker to craft harmful adversarial data points that are classified by the user as normal data points.
- Learners such as OCSVMs can withstand noise in data
- But are affected when adversaries deliberately distort data

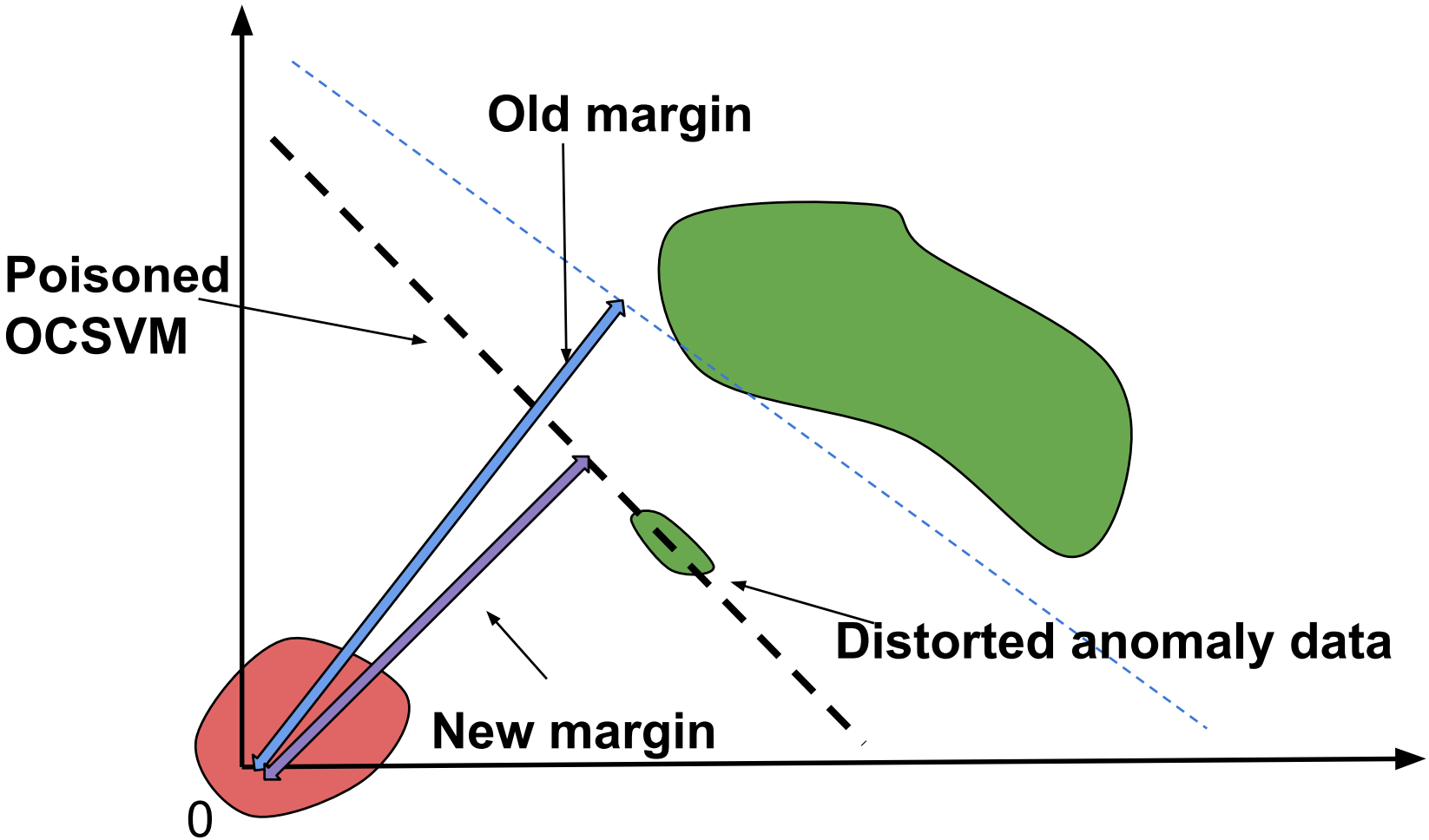
Increasing the Attack Resistance of OCSVMs

- There is a potential for adversarial distortions to have a less impact when data is projected to lower dimensions
- It becomes very difficult for the Adversary to predict the projection matrix because it is chosen randomly

OCSVM - BEFORE ATTACK



OCSVM - AFTER ATTACK



CONCLUSIONS

- OCSVMs are designed to withstand **noise** in training data
- But are vulnerable to malicious **adversarial distortions**
- Projecting training data to lower dimensional spaces could mask the possible adversarial distortions
- We have shown the difference can be reduced by projecting to lower dimensional spaces

P. Weerasinghe, S. Erfani, T Alpcan, C. Leckie, M. Kuijper, "Unsupervised Adversarial Anomaly Detection using One-Class Support Vector Machines," MTNS 2018.

Want more? MIT(Cyber Security)