INTERNET SECURITY

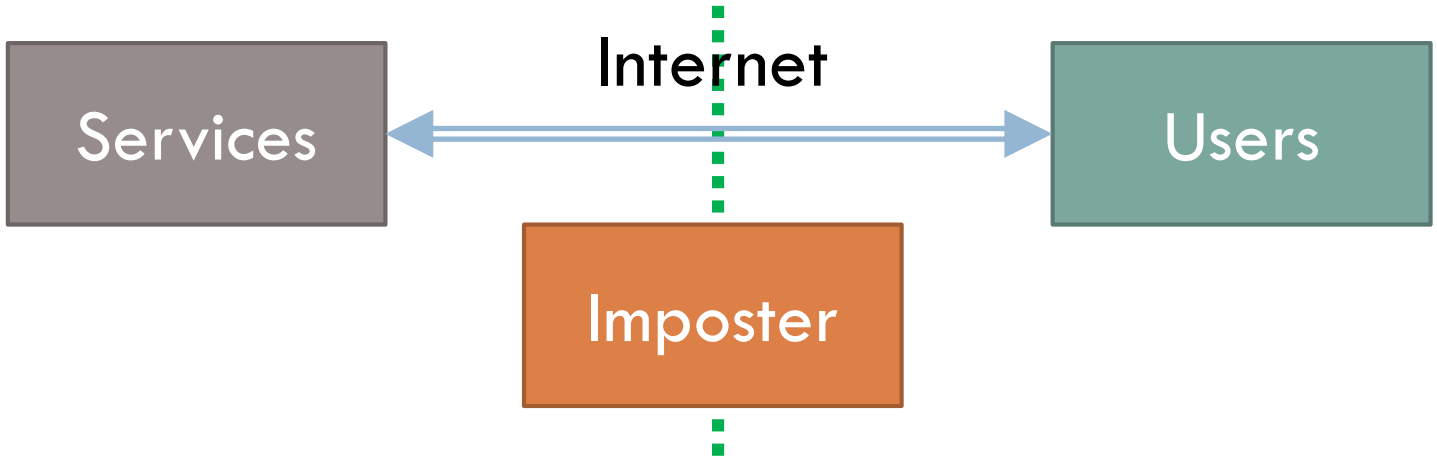
Based on materials Copter Labs

Overview

- ความปลอดภัยในเครือข่ายคอมพิวเตอร์ (Internet Security)
- รูปแบบการโจมตีจากแฮกเกอร์ (Hacker Attacks)

Internet Security

- ความปลอดภัยในเครือข่ายคอมพิวเตอร์ หมายถึงการรักษาข้อมูลที่ใช้ในการติดต่อระหว่างคอมพิวเตอร์ไม่ให้ถูกขโมย หรือหากข้อมูลถูกขโมย ผู้ขโมยไปต้องไม่สามารถถอดรหัสของข้อมูลนั้นได้
- การป้องกันไม่ให้ข้อมูลรั่วไหลสามารถทำได้สองวิธี
  - **Authenticate** การระบุตัวตนของผู้ที่รับข้อมูลนั้น
  - **Authorize** การให้อำนาจในการเข้าถึงข้อมูลของบุคคลในแต่ละระดับ



Authentication

- การระบุตัวตน คือ การตรวจสอบว่าบุคคลที่พยายามจะเข้ามาเรียกใช้ข้อมูลจะต้องเป็นตัวจริงและไม่ได้ถูกคนอื่นแอบอ้างมา
- เทคนิคที่ใช้ในการระบุตัวตนของผู้ใช้ข้อมูลแบ่งออกเป็นสามระดับ
  - ใช้สิ่งที่คุณมี (Possession)
  - ใช้สิ่งที่คุณรู้ (Memorization)
  - ใช้สิ่งที่คุณเป็น (Physicality)

Possession

- เราสามารถใช้สิ่งของที่ทำขึ้นมาเฉพาะเพื่อใช้ในการระบุตัวตนได้ เช่น
  - Smart Card
  - Student ID Card
  - Credit Card
- ข้อเสียของการใช้สิ่งของทำขึ้นมาเฉพาะ ได้แก่
  - ถ้าทำบัตรหาย ต้องไปทำใหม่ บัตรที่หายไปอาจจะถูกนำไปใช้ในทันทีได้
  - การปลอมแปลงบัตร เช่น ใช้เครื่องจดจำแถบแม่เหล็กซ่อนไว้ในแผงตู้ ATM
  - การยืมบัตรหรือขโมยบัตร

Memorization

- เราสามารถให้ผู้ใช้ข้อมูลจดจำรหัสผ่านเพื่อเข้าระบบก่อนที่จะสามารถดึงข้อมูลออกมาใช้ได้
- ข้อเสียของรหัสผ่าน ได้แก่
  - รหัสผ่านรั่วไหล ไม่ว่าจะเป็นการตั้งรหัสที่ง่ายเกินไป หรือผู้ใช้ข้อมูลเผลอบอกรหัสผ่านนี้ทั้งทางตรงและทางอ้อม
  - ผู้ประสงค์ร้ายสามารถดักจับรหัสผ่านได้ถ้าเข้าระบบโดยใช้คอมพิวเตอร์สาธารณะ
  - ระบบอาจจะถูกโจมตีด้วยโปรแกรมเดารหัสผ่าน
  - ถ้าหากลืมรหัสผ่านควรจะต้องทำอย่างไร

Physicality

- ใช้สิ่งที่เป็นเอกลักษณ์เฉพาะของแต่ละบุคคล (ขึ้นกับ DNA) เช่น
  - Face/Iris Recognition
  - Fingerprint/Palm Scan
  - Voice Authentication
- ข้อเสียของการใช้เอกลักษณ์เฉพาะ ได้แก่
  - ต้องใช้อุปกรณ์เฉพาะที่การตรวจจับ อาจจะไม่คุ้มกับการลงทุน
  - เทคโนโลยีสมัยใหม่สามารถทำการปลอมแปลงเอกลักษณ์ต่าง ๆ ได้ไม่ยาก
  - ผู้ใช้อาจจะโดนทำร้ายได้

Authorization

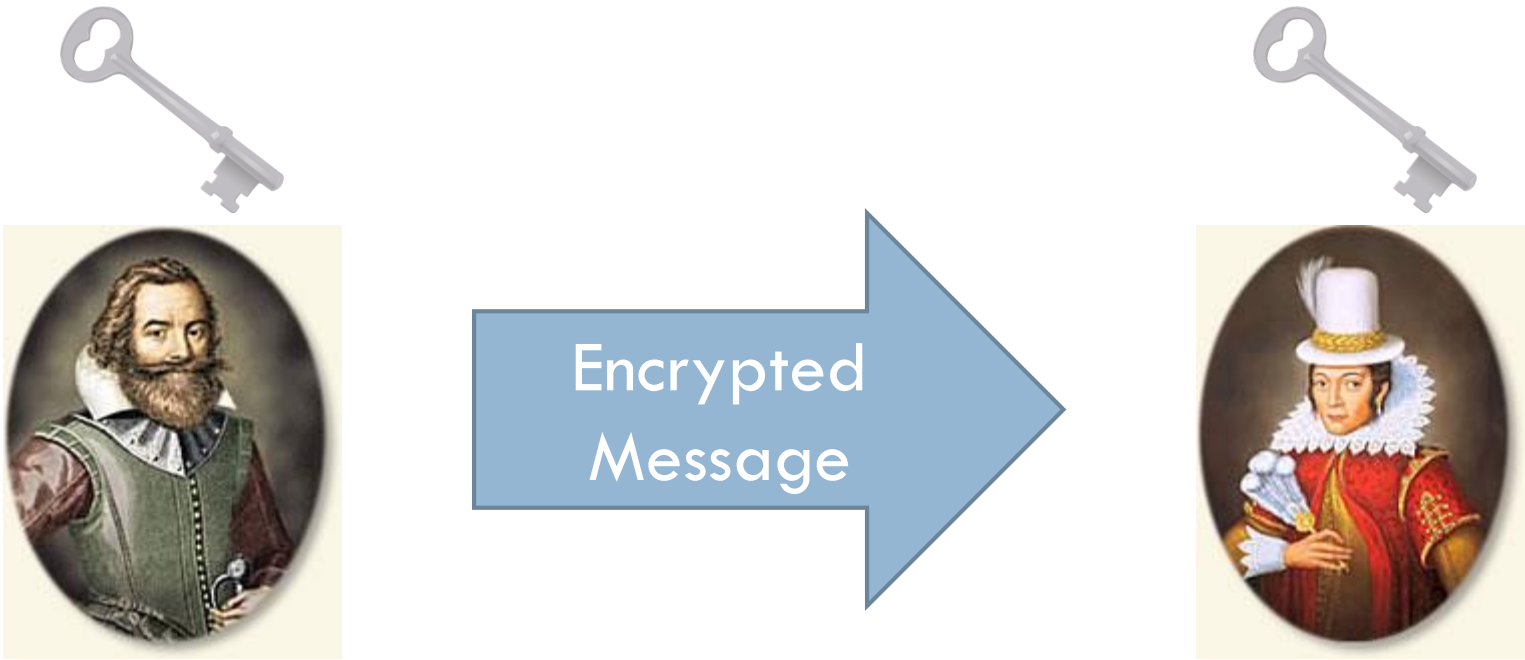
- การให้อำนาจในการเข้าถึงข้อมูลของบุคคลในแต่ละระดับมักจะคำนึงถึงความสำคัญหรือตำแหน่งของบุคคลนั้น ๆ เช่น
  - ผู้ใช้ A สามารถเห็นข้อมูลของตนเองได้แต่ไม่สามารถเห็นข้อมูลของผู้ใช้ B
  - ผู้ควบคุมระบบสามารถดูได้แค่ข้อมูลทั่วไปของผู้ใช้
- ปัญหาของการให้อำนาจแต่ละกลุ่มบุคคลได้แก่
  - ผู้ควบคุมระบบมีอำนาจในการเปลี่ยนแปลงรหัสผ่านของผู้ใช้ (อำนาจมากเกินไป)
  - ผู้ใช้ A ไม่สามารถเข้าไปแก้ไขชื่อของตนเองได้ (อำนาจไม่เพียงพอ)

Encryption and Decryption

- ในการส่งข้อมูลผ่านอินเตอร์เนต เราไม่สามารถรับประกันได้ว่าข้อมูลที่ส่งไปจะไม่ถูกดักระหว่างทาง
- ดังนั้นเราจึงจำเป็นต้องเข้ารหัสข้อมูล โดยข้อมูลที่ถูกเข้ารหัสแล้วนั้นสามารถถอดรหัสได้โดยผู้ส่งและผู้รับเท่านั้น
- การเข้ารหัสสามารถจำแนกได้สองระบบคือ
  - การเข้ารหัสโดยใช้คีย์ลับ (Secret Key Cryptography)
  - การเข้ารหัสโดยใช้คีย์สาธารณะ (Public Key Cryptography)

Secret Key Cryptography

- ใช้คีย์ลับในการเข้ารหัสและถอดรหัสข้อมูล ซึ่งคีย์ลับอาจจะตกลงกันไว้ก่อนการส่งข้อมูล เนื่องจากการส่งคีย์ลับผ่านอินเตอร์เนตอาจจะไม่ปลอดภัย



Transposition Ciphers

- เข้ารหัสข้อความด้วยการจัดเรียงใหม่

THIS IS A SECRET

เอาช่องว่างออก  แล้วจัดเรียงเป็นสามคอลัมน์

THI
SIS
ASE
CRE
T

อ่านจากบนลงล่าง  จากคอลัมน์ซ้ายไปขวา

TSACTHISRISEE

Substitute Ciphers

- เข้ารหัสข้อความด้วยเปลี่ยนตัวอักษรทีละตัว
  - เลื่อนตัวอักษรไปหนึ่งตัว

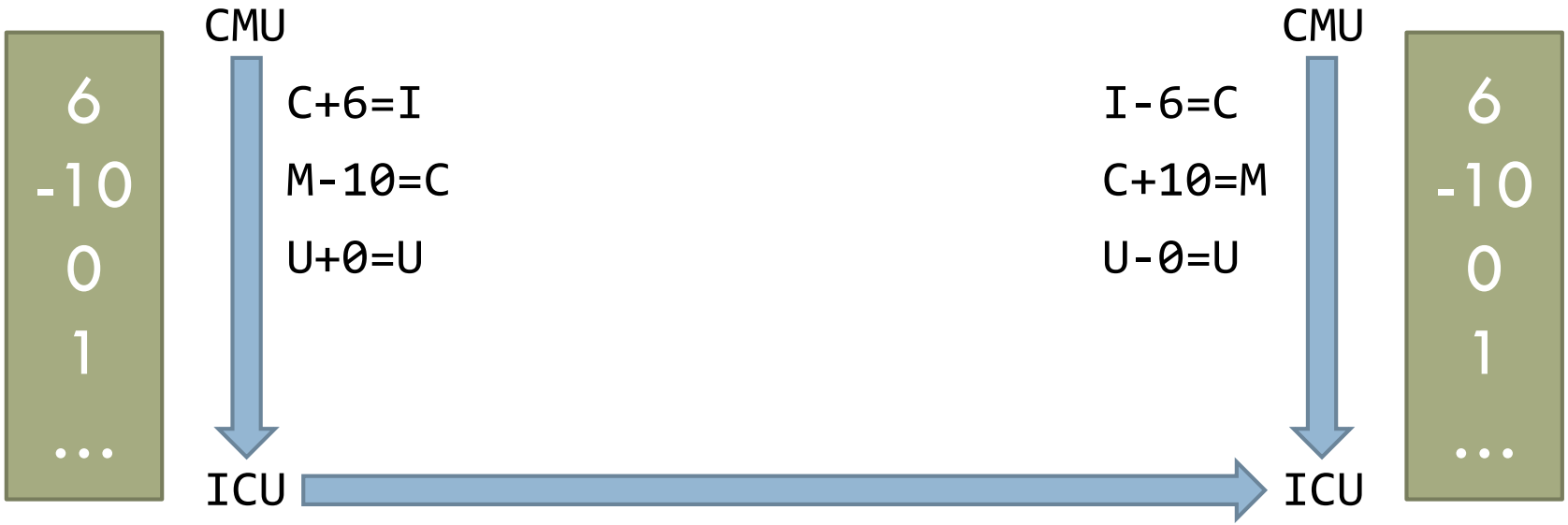
THIS IS A SECRET

UIJT JT B TFDSFU

- เราสามารถเดาข้อความที่เข้ารหัสในลักษณะนี้ได้หรือไม่
  - ดูคำที่มีจำนวนตัวอักษรน้อย ๆ

One Time Pads

- เข้ารหัสข้อความคล้ายกับวิธี **Substitute Cipher** เพียงแต่ว่าแต่ละตัวอักษรจะถูกเลื่อนไปเป็นจำนวนไม่เท่ากัน
- ผู้ส่งและผู้รับจะได้ตัวเลขที่สุ่มขึ้นมาชุดเดียวกัน โดยตัวเลขเหล่านี้จะเป็นตัวระบุว่าต้องเลื่อนตัวอักษรไปกี่ตำแหน่ง



Public Key Cryptography

- แต่ละฝ่ายของการรับส่งข้อมูลจะมีสองคีย์ คีย์สำหรับการเข้ารหัสและคีย์สำหรับการถอดรหัส
  - A มี e_A สำหรับเข้ารหัสและ d_A สำหรับถอดรหัส
  - B มี e_B สำหรับเข้ารหัสและ d_B สำหรับถอดรหัส
  - A ประกาศ e_A ให้สาธารณะรู้
  - B ส่งข้อความ M ให้ A โดยใช้ e_A ในการเข้ารหัส
  - A ถอดรหัสโดยใช้ d_A
- การเข้ารหัสและถอดรหัสจะใช้หลักการทางคณิตศาสตร์ในเรื่องการหารลงตัวและจำนวนเฉพาะ

RSA Keys Generation

- วิธีการ RSA เป็นวิธีการสร้างคีย์ e และ d เพื่อใช้ในการเข้าและถอดรหัส
- เริ่มต้นจากจำนวนเฉพาะขนาดใหญ่สองจำนวน p และ q
- กำหนดให้ n mod m คือเศษที่ได้จากการหาร n/m
- เลือก e และ d โดยที่
  - ed mod (p-1)(q-1) = 1
  - หรือเขียนในรูป e*d = 1 + k(p-1)(q-1)
  - e เป็นคีย์สาธารณะที่ใช้ในการเข้ารหัส
  - d เป็นคีย์ลับที่ใช้ในการถอดรหัส
  - pq เป็นข้อมูลสาธารณะ

RSA Public Keys Encryption

- เราสามารถใช้ **e** ในการเข้ารหัสได้ดังนี้
- แบ่งข้อความออกเป็นส่วนๆ จากนั้นแปลงแต่ละข้อความเป็นตัวเลข **m** โดยที่ $m < pq$
- เข้ารหัส **M** โดยการคำนวณค่า $m^e \bmod pq$
- การคำนวณค่านี้สามารถทำได้อย่างรวดเร็วด้วยวิธี **Modular Exponentiation**

RSA Private Keys Decryption

- ก่อนที่เราจะถอดรหัสได้ เราควรรู้จักสมการทางคณิตศาสตร์ที่น่าสนใจ
  - $m^{e+d} \bmod p = (m^e \bmod pq)*(m^d \bmod pq) \bmod pq$ (ง่าย)
  - $(m \bmod p)^d = m^d \bmod p$ (ง่าย)
  - $(m \bmod p) \bmod p = m \bmod p$ (ง่าย)
  - $m^{(p-1)(q-1)} \bmod pq = 1$ (ยาก)
- ถอดรหัสข้อความโดยการยกกำลัง **d** แล้ว **mod pq**
  - ทำไมทำแบบนี้แล้วจะได้ตัวเลขเดิมกลับมา

RSA Private Keys Decryption

- ถอดรหัสข้อความโดยการยกกำลัง d แล้ว mod pq

$$(m^e \bmod pq)^d \bmod pq = m^{ed} \bmod pq$$

$$= m^{1+k(p-1)(q-1)} \bmod pq$$

$$= (m \bmod pq)(m^{(p-1)(q-1)} \bmod pq)^k \bmod pq$$

$$= (m \bmod pq)(1)^k \bmod pq$$

$$= (m \bmod pq) \bmod pq$$

$$= m \bmod pq$$

$$= m$$ (เนื่องจาก $m < pq$)

Why RSA works?

- เพราะเหตุใดวิธีนี้จึงทำงานได้ดีและไม่สามารถเดาข้อความได้
- ค่า **d** มีขนาดใหญ่เกินกว่าจะเดาได้
- ถึงแม้จะรู้ค่า **e** และ **pq** เราไม่สามารถย้อนกลับมาคำนวณค่า **d** ได้
  - $ed \bmod (p-1)(q-1) = 1$
- หากเราต้องคำนวณค่าหา **p** และ **q** เราจำเป็นต้องลองจำนวนเฉพาะทุกคู่
  - ถ้า **pq** มีขนาด **664 bit** หรือ **200** หลัก ต้องใช้เวลา **3700** ปี
  - ถ้า **pq** มีขนาด **1024 bit** หรือ **308** หลัก ต้องใช้เวลา 10^{10} ปี
- ถ้าหากมีคนพิสูจน์ว่า $P = NP$ เป็นจริงแล้ว วิธีนี้อาจจะถูกเจาะได้ง่าย

Digital Signature

- คีย์สาธารณะ **e** และคีย์ลับ **d** สามารถทำให้คนอื่นไม่สามารถเข้าใจข้อความได้นอกจากผู้ส่งและผู้รับเท่านั้น
- นอกจากนี้แล้วคีย์ทั้งสองยังมีคุณสมบัติการสลับที่
  - **A** ส่งข้อความ **M** ให้ **B** โดยใช้ d_A ในการเข้ารหัส
  - **B** ถอดรหัสโดยใช้ e_A
- ดังนั้นการใช้คีย์ลับ **d** ในการเข้ารหัสถือเป็นลายเซ็นดิจิตอล (Digital Signature)

Central Key Distribution

- แต่เราจะทราบได้อย่างไรว่าคีย์สาธารณะที่ได้เป็นของอีกฝ่ายจริง
- เราสามารถแก้ไขปัญหานี้ได้โดยการมีคนกลางในการแจกจ่ายคีย์สาธารณะ
  - A และ B ส่ง e_A และ e_B ให้คนกลางตอนเข้าสู่ระบบครั้งแรกและรับคีย์สาธารณะของคนกลาง
  - A ร้องขอ e_B จากคนกลาง
  - คนกลางส่ง e_B ซึ่งจะถูกเซ็นดิจิตอลด้วยคีย์ลับของคนกลาง
  - A ใช้คีย์สาธารณะของคนกลางในการถอดรหัสเพื่อให้ได้ e_B กลับคืนมา

Public Key Infrastructure (PKI)

- Certificate Authorities (CA) เป็นองค์กรที่เชื่อถือได้และได้รับหน้าที่ในการออกใบรับรองหรือ Certificate ให้แก่ผู้ใช้และระบบหรือบริการที่ได้รับการตรวจสอบแล้ว
- ระบบ PKI สามารถนำมาใช้ในกระบวนการระบุตัวตนแบบอัตโนมัติได้ เนื่องจากใบรับรองจะถูกประทับตราออนไลน์ได้
- ใบรับรองจะประกอบไปด้วยคีย์สาธารณะของอีกฝ่ายซึ่งมีการเซ็นดิจิตอลแล้วโดย CA

Life is Tough

- ถึงแม้เราจะมีระบบที่ดูปลอดภัย คนบางกลุ่มยังสามารถเจาะระบบเข้ามาโดยผ่านช่องโหว่ของระบบ เช่น
  - ไวรัสและโปรแกรมที่ไม่พึงประสงค์
  - การโจมตีที่ทำให้ระบบทำงานเกินกว่าปกติ

Viruses

- ไวรัสคือโปรแกรมที่ทำงานบนคอมพิวเตอร์ที่ไม่ต้องการและไม่ได้รับอนุญาต
- มักจะปลอมแปลงหรือแฝงตัวเอง
- บางครั้งรอเวลาที่เหมาะสมก่อนจึงจะทำงาน
- ส่วนมากมักจะแก้ไขค่าต่าง ๆ ที่เกี่ยวข้องกับสิ่งแวดล้อมในเครื่อง
- เพราะเหตุใด **Windows** จึงติดไวรัสง่ายกว่าระบบปฏิบัติการอื่น
  - มีผู้ใช้ **Windows** จำนวนมากทำให้ความเสียหายเกิดในวงกว้าง
  - **OSX** ไม่ได้รันโปรแกรมนามสกุล **exe** ทำให้ผู้เขียนไวรัสต้องเขียนใหม่เอง
  - สำหรับ **Linux/UNIX** ผู้ใช้ไม่สามารถปรับเปลี่ยนสิ่งแวดล้อมในเครื่องได้ถ้าหากไม่มีการอนุญาตหรือเข้าระบบในฐานะ **root**

Antivirus

- ตรวจจับสัญญาณโดยสแกนไฟล์กับฐานข้อมูลไวรัสที่มีอยู่
- ต้องมีการอัพเดตฐานข้อมูลอยู่ตลอดเวลา เพราะมีไวรัสใหม่มาเรื่อย ๆ
- การตรวจสอบอาจจะเกิด **False Positive** ได้
- ไฟล์ที่โดนไวรัสอาจจะต้องถูกลบทิ้งถ้าหากต้องการลบไวรัสออกจากไฟล์

Denial of Service Attack

- ระดมส่งข้อความขยะจำนวนมากไปยังระบบเพื่อไม่ให้ข้อความที่มีประโยชน์เข้าถึงระบบ
  - ส่งอีเมลล์จำนวนล้านฉบับ
  - กด F5 ในหน้าเว็บรัว ๆ
- การโจมตีมีรูปแบบหลากหลาย
  - **Ping Flood** ส่งคำสั่ง **ping** จำนวนไปยังเหยื่อ ผู้ส่งจะต้องมี **bandwidth** ขนาดใหญ่กว่าเหยื่อถึงจะสำเร็จ
  - **R-U-Dead-Yet? (RUDY)** เป้าหมายจะเป็น **Web App** โดยพยายามใช้ **session** จนหมด โดยในแต่ละ **session** จะมีการส่งข้อมูลด้วยวิธี POST แบบไม่หยุด รวมทั้งส่ง **header** ขนาดใหญ่เพื่อใช้กินพื้นที่ **session**

Defense against DoS

- Firewall
  - ซ่อนโครงสร้างของเครือข่ายโดยทำให้ดูเหมือนว่าข้อมูลทั้งหมดออกมาจาก Firewall
  - ปิดกั้นข้อมูลจากผู้ใช้ที่ไม่ได้รับการยืนยันโดยการตรวจสอบที่อยู่หรือผู้ใช้ที่ไม่ได้อยู่ในรายชื่อที่เราสร้างไว้เอง
- Intrusion Detection System (IDS)
  - ใช้ Data Mining ในการตรวจจับและรายงานการกระทำที่น่าสงสัย โดยมีสองทางเลือกคือ Pattern Recognition และ Anomaly Detection