[보안 공학 졸업 고사 요령]

1. 일주일 전에 샘플 문제를 올려준다. (샘플 문제를 절대 믿으면 안 된다(100%변형됨) 그리고 검증 안 된 자료들도 절대 믿으면 안 된다. 지금 돌고 있는 답안은 틀린 것도 너무 많다^^;)
2. 문제는 영어로 출제 되며, 수업시간에 배웠던 부분과 과제에서 거의 다 나온다.
3. 용어 문제는 반드시 불특정 하게 나오며, 단어에 대한 정확한 뜻을 숙지하고 가야 함
4. 합격자도 자신이 왜 합격한지 모르는 시험, 불합격자는 더욱더 의문이 가는 결과를 맞게 됨. 무조건 토의하고 열심히 하 고 강의자료와 그 동안 나온 과제를 많이 풀어보는 방법이 유일하다. 절대 혼자만 풀면 망함.
5. 시간이 없으신 분 들은 [2011봄 졸업 고사 기출]을 제일 먼저 풀어 보시고 가세요.

[필독]

하나하나 검증해서 풀다 보니 시간이 오래 걸리네요..ㅜㅜ 시험 보기 전까지 100%가 완벽한 족보집이 되길 희망해 봅니다. 약 45% 정도 검증된 이 파일을 먼저 **제가 잘 아는 분께만 의견을 묻고자 베포** 합니다. **절대 다른 사람에게 주면 안됩니다**. **부탁 드립니다.** 아시다시피 완벽하지 않기 때문에 베포하면 저 욕먹어요..ㅜㅜ 문제에 대한 답이 이상하거나 완벽한 답을 알고 계시면 절대 혼자!! 보시지 말고 저한테 메일 보내주면 확인해보고 또 반영하도록 할게요. 지금 계속 수정 중입니다. shin830623@naver.com 제 메일주소입니다.

[자주 등장하는 필수 문제]

- 베이지안 스팸 필터
- 파레토법칙, 죄수딜레마, 내쉬 균형 문제
- L-diversity, Anomity 데이터베이스 프라이버시 보안
- Lamport's time stamp, 프로세스와 clock time 문제
- Lattice 와 Hasse diagrams 문제
- *-proporty, bibamodel, bel 라파듈라 모델 접근통제모델
- 기타 생일 역설 문제, 암호학 관련 문제, 온라인 광고 문제, 접근통제모델
- 단어 용어 정의 문제

1. 베이지안 스팸 필터

베이지안 스팸필터 문제를 풀기 전에 베이지안 이론에 대해서 알아보자. 베이지안 이론은 **과거에 A라는 사건이 발생했을 때, B라는 Class로 분류되었다면, 다음 번에도 A가 발생했을 때 B가 될 가능성이 높다는 의미**이다. 즉, **사전확률P(A)과 우도 확률(B|A)를 안다면 사후확률P(A|B)를 알 수 있다는 것**이다.

- 사전확률[prior probability] : A(원인)이 발생할 확률 P(A)와 같이 결과가 나타나기 전에 결정되어 있는 확률
- 우도확률(likelihood probability) : A(원인)이 발생하였다는 조건하에서 B(결과)가 발생할 확률P(B|A)을 나타낸다.

- 사후확률(posterior probability) : 사후 확률은 B(결과)가 발생하였다는 조건하에서 A(원인)이 발생하였을 확률을 나타낸다. 사후확률 즉 결과 B가 발생하였다는 조건 하에서 원인 A가 발생하였을 확률은 얼마인가?

$$P(A_i|B) = \frac{P(A_i \text{ and } B)}{P(B)}$$

$$P(A_i \text{ and } B) = P(B|A_i) \cdot P(A_i)$$

$$\begin{aligned} P(B) &= P(A_1 \text{ and } B) + \cdots + P(A_k \text{ and } B) \\ &= P(B|A_1)P(A_{1)} + \cdots + P(B|A_k)P(A_k) \end{aligned}$$

따라서 사후확률은 사전확률과 우도 확률들로 나타낼 수 있다.

$$\begin{aligned} P(A_i|B) &= \frac{P(A_i \text{ and } B)}{P(B)} \\ &= \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + \cdots + P(B|A_k)P(A_k)} \end{aligned}$$

[이해]

A : 개체가 가질 수 있는 어떤 성질

P(A) : 개체가 어떤 성질을 가질 확률(예: 키가 170이상일 확률, 메일이 어떤 단어를 포함할 확률)

Wi : class 전체 군을 몇 가지 성질에 따라 나눈 것이다.(예: 남자=W1, 여자=W2, 또는 스팸메일=W1, 정상메일=W2)

P(wi): 개체가 어떤 class에 속할 확률

P(Wi|A) = P(A|Wi)P(Wi)/P(A) : (posterior probability) P(Wi|A)는 개체가 어떤 성질 A를 가지고 있을 때 Class Wi에 속할 확률이다.

전체 군을 W1과 W2의 두 개의 class로 나누었다고 가정한다. 일단P(W1|A)+P(W2|A)=1이 된다. 어떤 개체는 W1이나 W2나 둘 중 하나의 군에는 반드시 속할 것이기 때문이다. 그렇다면, 지금까지의 경험(과거의 데이터)를 기준으로 P(W1|A)=0.8이고, P(W2|A)=0.2이라고 하자. 만일 미래에 A라는 성질을 가진 새로운 개체를 만난다면 우린 이 A를 W1에 분류가 가능할까? 과거의 데이터를 기준으로 우리는 A를 W1에 분류할 수 있을 것이다. 따라서 개체가 A라는 성질을 가졌을 때, P(Wi|A)값이 가장 큰 Wi에 A가 속할 것이라고 예측한다. P(W1|A)< P(W2|A)라고 하면, 어떤 개체가 A라는 성질을 가지면 W2라는 Class에 속한다고 예측할 수 있다. P(A|W1)P(W1)/P(A) < P(A|W2)P(W2)/P(A) 을 정리하면 P(A|W1)/P(A|W2) < P(W2)/ P(W1) (Bayes' decision rule) 이라고 할 수 있다. 베이지안 정리는 위의 식과 같다. 우리가 어떤 개체가 A라는 성질을 가지고 있다는 것을 알고 있을 때, 그 개체가 W1에 속할지 W2에 속할지 위의 식으로 알 수 있다. 위의 식에서 왼쪽이 더 크다면 classW1에 속할 것이고, 오른쪽이 더 크다면 classW2에 속한다. 왼쪽 값, 오른쪽 값 모두 '경험'을 통해서 얻어낼 수 있는 값이다. 전문적인 용어로 왼쪽은 likelihood ratio 오른쪽은 threshold라고 한다. 즉, ratio가 임계 값을 넘느냐 안 넘느냐에 따라서 class가 결정된다.

Ex) 우리가 받는 메일의 80%가 스팸 메일이다. W2를 spam메일 class, W1을 정상 메일 class 라고

하면 오른쪽 threshold 값은 0.8/0.2 = 4 라고 할 수 있다. 그럼 왼쪽 값이 4보다 작으면 spam메일, 4보다 크면 정상메일 이라고 규정할 수 있다.

[반드시 암기 공식]

P(A|B)= $\frac{P(A\cap B)}{P(B)}$, P(B|A)= $\frac{P(A\cap B)}{P(A)}$, P(A∩B) = P(B|A)*P(A) , P(A|B)= $\frac{P(B|A)P(A)}{P(B)}$

[2011봄 졸업고사 기출]

Suppose that a Bayesian spam filter is trained on a set of 10,000 spam messages and 4,000 messages that are ham. The word "vaccine" appears in 1,500 spam messages and 20 messages that are ham, while the word "virus" appears in 800 spam messages and 200 messages that are ham. Estimate the probability that a received message containing both the words "vaccine" and "virus" is spam. Will the message be rejected as spam if the threshold for rejected as spam if the threshold for rejecting spam is 0.9? (20point)

What is the Bayes theorem? Explain the relationship between the conditional probability P(A|B) and P(B|A) and the marginal probability P(A) and P(B).

베이지정리란 사전확률(prior probability)를 이용하여 사후확률(posterior probability)을 추정하는 데에 활용되는 정리

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}=\frac{P(B|A)P(A)}{P(A)P(B|A)+P(\bar{A})P(B|\bar{A})}$$

여기서, 사전 확률 : P(A), P(B|A), $P(\bar{A})$, $P(B|\bar{A})$

사후 확률 : P(A|B)

$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{P(B|A)P(A)}{P(B)}$ ➔ B가 일어났을 때 A가 일어날 확률

$P(B|A)=\frac{P(A\cap B)}{P(A)}=\frac{P(A|B)P(B)}{P(A)}$ ➔ A가 일어났을 때 B가 일어날 확률

A가 일어났을 때 B가 일어날 확률은 쉽다. 하지만 B가 일어났을 때 A가 일어날 확률은 상당히 어렵다. 따라서 우리는 쉽게 구할 수 있는 $P(A|B)$ 를 통해서 $P(B|A)$ 를 구하는 것이다.

A. Express the marginal probability P(B) by using. The law of total probability.

$$P(B)=P(B\cap A)+P(B\cap\bar{A})=P(B|A)P(A)+P(B|\bar{A})P(\bar{A})$$

B. Complete the spam filtering formula based on the Bayes theorem and the law of total probability.

$$P(A|B)=\frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|\bar{A})P(\bar{A})}$$

C. What is your conclusion about the spam filtering?

$P(S|A\&B) = \frac{P(A\&B|S)P(S)}{P(A\&B|S)P(S)+P(A\&B|H)P(H)}$ = A와 B가 동시에 스팸일 확률

$$= \frac{P(A|S)P(B|S)P(S)}{P(A|S)P(B|S)P(S) + P(A|H)P(B|H)P(H)}$$

where S denotes spam, H ham, A vaccine, and B virus. For naïve filter, $P(H) = P(S) = 0.5$. *Thus,*

$$P(S|A\&B) = \frac{P(A|S)P(B|S)}{P(A|S)P(B|S) + P(A|H)P(B|H)}$$

- 10,000 개의 스팸 메시지
- 4,000 개의 정상 메시지
- 1,500 개의 스팸 메시지에 vaccine
- 20 개의 정상 메시지에 vaccine
- 800 개의 스팸 메시지에 virus
- 200 개의 정상 메시지에 virus

$$P(S) = \frac{10000}{10000 + 4000} = \frac{10}{14}$$

$$P(H) = \frac{4000}{10000 + 4000} = \frac{4}{14}$$

$$P(A|S) = \frac{1500}{10000} = 0.15$$

$$\mathrm{P(A|H)} = \frac{20}{4000} = 0.005$$

$$\mathrm{P(B|S)} = \frac{800}{10000} = 0.08$$

$$\mathrm{P(B|H)} = \frac{200}{4000} = 0.05$$

1) *When* $P(S) = P(H) = 0.5$,

$$P(S|A\&B) = \frac{P(A|S)P(B|S)}{P(A|S)P(B|S) + P(A|H)P(B|H)}$$
$$= \frac{0.15 * 0.08}{0.15 * 0.08 + 0.005 * 0.05}$$
$$= 0.9796$$

2) *When* $P(S) = \frac{10}{14}, P(H) = \frac{4}{14}$

$$P(S|A\&B) = \frac{P(A|S)P(B|S)P(S)}{P(A|S)P(B|S)P(S) + P(A|H)P(B|H)P(H)}$$
$$= \frac{0.15 * 0.08 * \frac{10}{14}}{0.15 * 0.08 * \frac{10}{14} + 0.005 * 0.05 * \frac{4}{14}}$$

$$= 0.9917$$

위 결과는 spam으로 거부되는 임계값 0.9보다 크기 때문에 이 메시지는 spam 처리 되어 거부된다.

[2011 기말고사 기출]

어떤 Bayesian spam filter를 훈련시키는데 10,000개의 spam message와 4,000개의 ham message를 사용한다. 그런데 spam message에는 virus라는 단어가 500개의 spam message에 포함되어 있고 ham message에서는 200개에서 나온다. 수신한 message에 virus라는 단어가 포함되어 있을 때 그 message가 spam일 확률은 얼마인가? 그리고 Bayes' theorem을 사용하는 이유를 기술하라? (10점)

- 10,000 개의 스팸 메시지
- 4,000 개의 정상 메시지
- 500 개의 스팸 메시지에 virus
- 200 개의 정상 메시지에 virus

$$P(H) = \frac{4000}{10000 + 4000} = \frac{2}{7}$$

$$P(S) = \frac{10000}{10000 + 4000} = \frac{5}{7}$$

$$P(V|H) = \frac{200}{4000} = \frac{1}{20}$$

$$P(V|S) = \frac{500}{10000} = \frac{1}{20}$$

$P(H) = P(S) = 0.5$라고 가정할 경우

$$P(S|V) = \frac{P(V|S)}{P(V|S) + P(V|H)} = \frac{\frac{1}{20}}{\frac{1}{20} + \frac{1}{20}} = 0.5$$

$P(H) \neq P(S) = 0.5$라고 가정할 경우

$$P(S|V) = \frac{P(V|S)P(S)}{P(V|S)P(S) + P(V|H)P(H)} = \frac{\frac{1}{20} \cdot \frac{5}{7}}{\frac{1}{20} \cdot \frac{5}{7} + \frac{1}{20} \cdot \frac{2}{7}} = 0.71$$

Bayes 정리를 사용하는 이유는 $P(V|S)$ 나 $P(V|H)$ 은 상대적으로 쉽게 계산할 수 있으나 $P(S|V)$은 계산이 쉽지 않으므로 계산하기 쉬운 것을 이용해서 어려운 계산을 수행하기 위해 사용한다.

[2011 2학기 homework]

Suppose that a Bayesian spam filter is trained on a set of 11,000 spam messages and 4,000 messages that are ham. The word "improved" appears in 1,500 spam messages and 20 messages that are ham, while the word "aloe" appears in 800 spam messages and 200 messages that are ham. Estimate the probability that a received message containing both the words "improved" and "aloe" is spam. Will the message be rejected as spam if the threshold for rejected as spam if the threshold for rejecting spam is 0.9?

- 11,000 개의 스팸 메시지
- 4,000 개의 정상 메시지
- 1,500 개의 스팸 메시지에 "improved"
- 20 개의 정상 메시지에 "imporved"
- 800 개의 스팸 메시지에 "aloe"
- 200 개의 정상 메시지에 "aloe"

$$P(S|imporved\ \&aloe) = \frac{P(imporved\ \&aloe|S)P(S)}{P(imporved\ \&aloe)}$$
$$= \frac{P(imporved\ \&aloe|S)P(S)}{P(imporved\ \&aloe|S)P(S) + P(imporved\ \&aloe|H)P(H)}$$
$$= \frac{P(imporved|S)P(aloe|S)P(S)}{P(imporved|S)P(aloe|S)P(S) + P(imporved|H)P(aloe|H)P(H)}$$

(improved 와 aloe 는 서로 독립시행)

여기서 $P(improved|S) = \frac{1500}{11000}$, $P(improved|H) = \frac{20}{4000}$,
$P(aloe|S) = \frac{800}{11000}$, $P(aloe|H) = \frac{200}{4000}$ 가 되나

스팸과 햄의 비율을 명확히 언급하지 않았으므로 이 비율은

1) $P(S) = P(H) = 0.5$ 라고 볼 수도 있고,

2) $P(S) = \frac{11000}{15000}$, $P(H) = \frac{4000}{15000}$ 으로 볼 수도 있다.

3) $P(S) = P(H) = 0.5$ 라고 가정하면 이 항들이 약분되어

$$P(S|imporved\ \&aloe) = \frac{\frac{1500}{11000} * \frac{800}{11000}}{\frac{1500}{11000} * \frac{800}{11000} + \frac{20}{4000} * \frac{200}{4000}}$$
$$= 0.9754$$

4) $P(S) = \frac{11000}{15000}, P(H) = \frac{4000}{15000}$ 라고 가정하면

$$P(S|imporved\ \&aloe) = \frac{\frac{1500}{11000} * \frac{800}{11000} * \frac{11000}{15000}}{\frac{1500}{11000} * \frac{800}{11000} * \frac{11000}{15000} + \frac{20}{4000} * \frac{200}{4000} * \frac{4000}{15000}}$$
$$= 0.9909$$

[2010 기말고사 기출]

Suppose that a Bayesian spam filter is trained on a set of 10,000 spam message and 5,000 messages that are ham. The word "improved" appears in 1,500 spam messages and 10 messages that are ham, while the word "aloe" appears in 800 spam messages and 200 messages that are ham.

A. Estimate the probability that a received message containing the word "improved" is spam. Will the message be rejected as spam if the threshold for rejected as spam if the threshold for rejecting spam is 0.9 [10 points]

- 10,000 개의 스팸 메시지
- 5,000 개의 정상 메시지
- 1,500 개의 스팸 메시지에 "improved"
- 10 개의 정상 메시지에 "imporved"
- 800 개의 스팸 메시지에 "aloe"
- 200 개의 정상 메시지에 "aloe"

$$P(H) = \frac{5000}{10000 + 5000} = \frac{1}{3}$$

$$P(S) = \frac{10000}{10000 + 5000} = \frac{2}{3}$$

$$P(in\ proved\ \ |H) = \frac{10}{5000} = \frac{1}{500}$$

$$P(in\ proved\ \ |S) = \frac{1500}{10000} = \frac{3}{20}$$

$P(H) = P(S) = 0.5$라고 가정할 경우

$$P(S|in\ proved\ \ ) = \frac{P(in\ proved\ \ |S)}{P(in\ proved\ \ |S) + P(in\ proved\ \ |H)} = \frac{\frac{3}{20}}{\frac{3}{20} + \frac{1}{500}} = 0.152$$

$P(H) \neq P(S) = 0.5$라고 가정할 경우

$$P(S|in\ proved\ \ ) = \frac{P(in\ proved\ \ |S)P(S)}{P(in\ proved\ \ |S)P(S) + P(in\ proved\ \ |H)P(H)} = \frac{\frac{3}{20} \cdot \frac{2}{3}}{\frac{3}{20} \cdot \frac{2}{3} + \frac{1}{500} \cdot \frac{1}{3}}$$
$$= 0.993377$$

위 결과는 spam으로 거부되는 임계값 0.9보다 크기 때문에 이 메시지는 spam 처리 되어 거부된다.

B. Estimate the probability that a received message containing both the keywords "improved" and "aloe" is spam. Will the message be rejected as spam if the

threshold for rejected as spam if the threshold for rejecting spam is 0.9? [10 points]

- 10,000 개의 스팸 메시지
- 5,000 개의 정상 메시지
- 1,500 개의 스팸 메시지에 ”improved”
- 10 개의 정상 메시지에 “imporved”
- 800 개의 스팸 메시지에 “aloe”
- 200 개의 정상 메시지에 “aloe”

$$P(S) = \frac{10000}{10000 + 5000} = \frac{2}{3}$$

$$P(H) = \frac{5000}{10000 + 5000} = \frac{1}{3}$$

$$\mathrm{P(I|S)} = \frac{1500}{10000} = 0.15$$

$$\mathrm{P(I|H)} = \frac{10}{5000} = 0.002$$

$$\mathrm{P(A|S)} = \frac{800}{10000} = 0.08$$

$$\mathrm{P(A|H)} = \frac{200}{5000} = 0.04$$

1) *When* $P(S) = P(H) = 0.5$,

$$P(S|A\&I) = \frac{P(A|S)P(I|S)}{P(A|S)P(I|S) + P(A|H)P(I|H)} = \frac{0.08 * 0.15}{0.08 * 0.15 + 0.04 * 0.002} = 0.99337$$

2) *When* $P(S) = \frac{2}{3}, P(H) = \frac{1}{3}$

$$P(S|A\&I) = \frac{P(A|S)P(I|S)P(S)}{P(A|S)P(I|S)P(S) + P(A|H)P(I|H)P(H)} = \frac{0.08 * 0.15 * \frac{2}{3}}{0.15 * 0.08 * \frac{2}{3} + 0.005 * 0.05 * \frac{1}{3}} = 0.98984$$

위 결과는 spam으로 거부되는 임계값 0.9보다 크기 때문에 이 메시지는 spam 처리 되어 거부된다.

2. Pareto Law, 내쉬 균형 문제

파레토 개선(Pareto improvement) : 하나의 자원배분 상태에서 다른 사람에게 손해가 가지 않게 하면서 최소한 한 사람 이상에게 이득을 가져다 주는 것을 말한다.
파레토 최적(Pareto optimality) & 파레토 효율(efficiency) : 자원배분이 가장 효율적으로 이루어진 상태를 말함. **즉 파레토 개선이 불가능한 상태**, 즉, 다른 사람에게 손해가 가도록 하지 않고서는 어떤 한 사람에게는 이득이 되는 변화를 만들어내는 것이 불가능할 때 이 배분 상태를 이른다.

Ex) 현재 아르바이트생 A와 B가 있고 그들은 똑같이 사업에 기여를 했다고 한다. 그들에게 줄 수 있는 **돈은 총 1만원**(하나의 자원)이다. 그리고 그들에게 주고 남은 돈은 무조건 버릴 수 밖에 없다고 가정한다. 어떻게 배분해 줄까 고민하다가, **A에게 5천원, B에게 3천원을 줄 수 있다. 이는 파레토 비효율적** 이라고 한다. 남는 **2천원을 배분해주면(즉, 파레토 개선**을 하면)누군가의 효용이 증가 하기 때문이다. 예를 계속 들면,

1. A에게 5천원, B에게 5천원을 줄 수 있다. 이는 **파레토 효율적**이다.
2. A에게 7천원, B에게 3천원을 주어도 이는 **파레토 효율적**이다.
3. A에게 만원 전부 다, B에게 0원을 주어도 이는 **파레토 효율적**이다.

내쉬 균형 이론(Nash equilibrium) : 게임이론에서 가장 많이 사용되는 평형 개념이다. 상대적으로 모두 만족하고 있는 상태를 말한다. 모든 사람이 상대의 전략에 대해 최대 효용을 갖는 전략을 선택한 상태이다. 기존 경제학에서는 사람들이 자신의 욕망을 채우기 위해서 이기적으로 행동하고 그런 결과로 사회전체의 부가 증대된다고 보았다. 하지만 내쉬는 만약 모든 사람이 자신만의 욕망을 채우기 위해 상대방을 고려하지 않고 이기적으로 행동한다면 모두 실패하게 될 것이라고 생각하였다. 내쉬는 상대방을 고려하면서 자신의 만족이 최대가 되도록 행동한다면 모두가 최대의 만족을 가질 수 있다는 것을 보여주었다.

Ex) 내쉬 평형이 깨짐은.. 만약 파티에서 모든 남자들이 가장 예쁜 여자에게만 몰리면 오직 한 남자만이 그 여자와 즐거운 시간을 보낼 수 있게 되고 나머지 여자들은 자존심이 상해 다른 남자들을 만나주지 않게 되므로 나머지 남자들 여자들은 즐거운 파티를 즐길 수 없게 된다.

죄수의 딜레마(Prisoners' Dilemma) : 두 공범자가 서로 협력해 범죄사실을 숨기면 증거 불충분으로 형량이 낮아지는 최선의 결과를 누릴 수 있음에도 불구하고, 상대방의 범죄사실을 밝혀주면 형량을 감해 준다는 수사관의 유혹에 빠져 상대방의 죄를 고변함으로써 무거운 형량을 선고 받게 되는 현상을 말한다.

우월전략(dominant strategy) : 상대방이 어떤 전략을 선택하는지 관계 없이 자신의 보수를 더 크

게 만드는 전략을 우월전략이라 하며, 그 짝을 우월전략 균형이라고 부른다.

[2011봄 졸업고사 기출]

Regarding the following payoff table, answer the questions: (20 points)

		Player2	
		Collude	Confess
Player 1	Collude	x, y	1,10
	Confess	10,1	5,5

A. Explain what Pareto improvement and Pareto optimality are.

Pareto improvement이란 하나의 자원배분 상태에서 다른 사람에게 손해가 가지 않게 하면서 최소한 한 사람 이상에게 이득을 가져다 주는 것을 말함
Pareto efficient 또는 Pareto optimal 이란 Pareto improvement(개선)이 불가능한 상태 자원의 배분이 가장 효율적으로 이루어진 상태를 말함

B. Explain what Nash equilibrium is.

*The **Nash equilibrium** exists if no players change their strategy, despite knowing the actions of their opponents. In other words, in Nash equilibrium, no player has an incentive to deviate from the strategy chosen, since no player can choose a better strategy given the choices of the other players.*

내쉬 균형이란? 참여자가 어떤 특정한 전략을 선택해서 하나의 결론에 도달했을 때, 모든 참여자가 이에 만족하고 자신의 선택이 최선이라고 여기며 더 이상 전략을 변화시킬 의도가 없는 경우를 '내쉬 균형'에 도달함
내쉬 균형은 각자가 최적의 결과를 예상하고 행한 행동의 종합이지만, 이로부터 반드시 최적의 결과가 나옴을 보장하는 것은 아니다. 아니, 오히려 나쁜 결과를 포함하는 것이야말로 내시 균형이라고 부를 수 있다. 상대방의 최적전략을 예상하고 수립한 최적전략이지만 상대방이 그 최적전략을 내놓지 않으면 자신의 전략도 최적이 아니게 된다는 뜻이다

(1) 경매에서 각각의 입찰자가 얼마에 입찰할지는 타인의 입찰 예상액에 따른다. 이때 모든 입찰자가 합리적으로 생각해서 자신의 입찰가를 결정한 결과가 내시 균형이다.

(2) 두 남녀가 데이트를 하는데 발레 공연과 권투 경기가 겹쳤다. 남자는 권투를, 여자는 발레를 더 좋아하지만 각자 좋아하는 것을 혼자 보는 것보다는 같이 보는 것을 좋아한다. 이때 상대방의 행동을 예상해서 각자가 무엇을 볼지 결정한 결과가 내시 균형이다.

C. Explain what prisoners' dilemma is.

A game where each player has two strategies, say collude or confess, such that for each player, confess dominates collude, and the outcome (confess, confess) is worse for both than the outcome (collude, collude).

정의]

두 공범자가 서로 협력해 범죄사실을 숨기면 증거 불충분으로 형량이 낮아지는 최선의 결과를 누릴 수 있음에도 불구하고, 상대방의 범죄사실을 밝혀주면 형량을 감해 준다는 수사관의 유혹에 빠져 상대방의 죄를 고변함으로써 무거운 형량을 선고 받게 되는 현상을 말한다.

이해]

두 명의 공범자 1, 1를 형사가 검거했다. 그러나 그 형사는 심증만 있는 상태라서 범인들의 자백이 없이는 기소가 불가능하다. 그러자 이 형사는 다음과 같은 방법을 생각해 냈다. 각각 피의자를 독방에 격리시켜 놓고 다음과 같은 사항을 알려 주었다. "만약 **당신이 자백을 하지 않았는데 다른 한 명이 자백을 하면 자백한 사람은 특전으로 풀려나나** 당신은 **10년의 징역**을 살게 됩니다. 반대로 당신만이 자백을 하게 되면 당신이 특전의 혜택을 받게 됩니다. **만약 둘이 다 자백하는 경우에는 각각 5년 형을 언도 받을 것이며**, **둘 다 자백을 하지 않으면 각각 1년 형을 언도 받게 될 것입니다."**

위 상황에서 **1과 2는 각각 자백하는 것이 Domanant Strategy이다. 따라서 (C,C)는 Dominant Strategy Equilibrium이 된다. 그 결과 두 용의자는 모두 5년씩 징역을 살게 된다.** 이 결과를 살펴보면, 서로에게 불만족한 해를 얻는다는 문제를 발견할 수 있는데, 그것은 둘이 다 자백하면 1년씩만 징역을 살면 되는데 두 사람에게 서로 좋은 결과를 마다하고 5년 징역을 살게 되는 것이 최선의 대응이라는 결과가 나오는 까닭에 딜레마이다.

D. For what values of x, will this game be a prisoners' dilemma?

The condition to be a Prisoners' dilemma is as follows:

$$5 < x, y < 10.$$

E. Choose the appropriate values of x and y so that the solution has at least one Nash equilibrium which is also Pareto efficient.

$$10 \leq x, y \text{ or } x, y \leq 5.$$

[2010 중간고사 기출]

If two spiders find a dead insect at the same time, each spider will make menacing gestures to scare off the other. If one spider backs down, that spider gets nothing and the other spider gets the insect to itself. If both spiders back down, they can share the insect. If neither backs down, the spiders will fight. The payoffs resulting from the fight depend on the sizes of the spiders and are described below.

		Spider2	
		Back down	Fight
Spider 1	Back down	5,5	0,10
	Fight	10,0	x,y

A. Suppose the spiders are the same size so that x =y. For what values of x, will all strategies be Pareto efficient? (5points)

B. Suppose the spiders are the same size. For what values of x, will this game be a Prisoners' Dilemma? What is the Nash equilibrium Strategy? Is this Nash equilibrium Pareto efficient? (5points)

C. Suppose the spiders are the same size. For what values of x, will this game have one Nash equilibrium strategy that is also Pareto efficient? (5points)

[2010 중간고사 샘플 문제]

If 2 spiders find a dead insect at the same time, each spider will make menacing gestures to scare off the other. If one spider backs down, that spider gets nothing and the other spider gets the insect to itself. If both spiders back down, they can share the insect. If neither backs down, the spiders will fight. The payoffs resulting from the fight depend on the sizes of the spiders and are described below.

		Spider 2	
		Back down	Fight
Spider 1	Back down	5, 5	0, 10
	Fight	10, 0	x, y

A. Suppose the spiders are the same size so that $x = y$. For what values of x, will each spider have dominant strategy? What is the dominant strategy? (Show your

work)

Spider1이 Back down해서 얻을 수 있는 보상은 5나 0이며, Fight해서 얻을 수 있는 보상은 10과 X이다. Spider2 또한 같은 보상을 가지게 된다. Fight, Fight에서 얻을 수 있는 x=y가 0보다 크기만 하면, 두 Spider는 우세 전략을 가지게 된다. 따라서 두 거미가 우세 전략을 가질 조건은 x,y 둘 다 0보다 크기만 하면 된다.

1)정의

상대방이 어떤 전략을 선택하는지 관계없이 자신의 보수를 더 크게 만드는 전략을 우월전략이라 하며, 그 짝을 우월전략균형이라고 부른다.

2)성질

① 우월전략이 존재한다면 유일하며, 안정적이다.
② 효율성을 담보하지는 않는다.
③ 우월전략균형은 내쉬균형이 된다.

B. Suppose the spiders are the same size. For what values of x, will this game be a Prisoners' Dilemma? (Show your work)

죄수의 딜레마는 합리적인 두 죄수가 만약 둘 중 하나라도 부인을 택했다면, 더 많은 보상을 받을 수 있음에도 불구하고, **자신들의 우세전략에 따라 부인 대신 고백을 선택한 상황을 말한다**. 이 경우에도 두 거미가 열등전략, 여기서는 물러남의 보상이 큼에도 불구하고 우세전략, 여기서 싸움을 택하게 된다면, 죄수의 딜레마에 쳐하게 된다. 따라서 우세 전략의 보상자체가 열등전략 보상보다 작아야지만, 죄수의 딜레마 경우를 가지기 때문에 x,y는 0보다 크고 5보다 작아야 한다. 답 : $0 < x=y < 5$

[2011 2학기 기말고사 기출]

Use the payoff matrix given below and show what the Nash equilibrium is, if any, for both pure and mixed strategies. (10pt)

	Left	Right
Up	2,0	0,0
down	0,0	0,2

3. K-Anomity , L-diversity 데이터베이스 프라이버시 보안

데이터 익명화를 위한 연구에서는 테이블로 구성된 관계형 데이터(relational data)의 속성을 네가지로 분류한다. 각 레코드(record)가 어떤 개인의 정보를 포함하는 지 가르키는 **식별자(identifier)**, 속성들의 조합으로 개인을 특징지어주는 **준식별자(quasi-identifier)**, 개인에게 밀

접한 정보로 보호되어야 하는 **민감한 속성(sensitive attribute)**, 앞의 세가지 부류에 속하지 않는 나머지 속성으로 나누어진다.

K-Anomity – 전체 데이터 내에서 동일한 준 식별자를 갖는 레코드들의 수가 적어도 K개 이상이 되는 프라이버시 모델이다. 이 모델을 만족하는 데이터는 준 식별자로 특정 레코드를 유일하게 판별하지 못하게 만들어 프라이버시 침해 위험을 줄인다. 하지만 k-anonymity로는 민감한 속성을 파악하는 속성 결합(attribute linkage)공격이 발생할 수 있다. k-anonymity 익명성을 제공하기는 하지만 동등 클래스 내에서 민감한 속성의 값이 모두 같다면 공격자는 개인의 민감한 속성을 정확하게 알아낼 수 있다. L-diversity는 이러한 문제점을 해결하기 위해 동등 클래스 안에서 가장 빈번하게 발생하는 민감한 속성의 확률이 최대 1/1이 되도록 하여 이러한 문제점을 해결한다. K나 L값이 커질수록 프라이버시 보호 정도도 증가한다. 그러나 일반화로 인한 정보손실로 데이터의 유용성은 감소한다.

L-diversity – 데이터를 가공하여 준 식별자를 통해 민감한 속성을 추론할 확률을 1/L이하가 되는 프라이버시 모델이다. 즉, 구분되지 않은 레코드들로 이루어진 각 집합의 민감한 속성들은 적어도 1개의 서로 구분되는 민감한 속성들을 포함하도록 하는 기법이다.

[2011봄 졸업고사 기출]

Regarding the privacy enhancement technology, answer the following questions. (20 points)

A. What are *k*-anonymity and *l*-diversity?

A table satisfies k-anonymity if every record in the table is indistinguishable from at least k − 1 other records with respect to every set of quasi-identifier attributes. The main idea behind ℓ-diversity is the requirement that the values of the sensitive attributes are well-represented in each group.

B. Redesign the following database based on 3-anonymity principle as best as you can.

Name	Date of Birth	ZIP code	Sex	Blood	Illness
Kennedy	1983	12335	F	A	Breast cancer
Einstein	1984	14567	M	B	HIV
Lee	1960	12346	M	O	Manic-depressive
Hanks	1963	51234	F	AB	Arteriosclerosis
Hanks	1965	51236	F	O	Angina pectoris
Wang	1968	51234	M	O	Colitis

Aitken	1976	51237	F	O	Mellitus
Taroni	1960	51236	M	A	Gastric hyperacidity
Tsomko	1954	14567	M	B	Empyema
Kim	1998	12347	F	O	Epilepsy
Yang	2000	51235	M	O	Hypotension
Tanaka	2001	14567	M	B	Precocious dementia

C

Date of Birth	ZIP code	Sex	Blood	Illness
198*	1234*	F	A	Breast cancer
198*	1234*	F	O	Epilepsy
198*	1234*	F	O	Manic-depressive
196*	5123*	*	A	Gastric hyperacidity
196*	5123*	*	O	Angina pectoris
196*	5123*	*	O	Colitis
2000	5123*	F	O	Hypotension
2000	5123*	F	AB	Arteriosclerosis
2000	5123*	F	O	Mellitus
19**	14567	*	B	Empyema
19**	14567	*	B	Precocious dementia
19**	14567	*	B	HIV

D. Why a cell with just two entries (see the Chemistry-Geology pair) should be suppressed?

Assume that one of the two students taking Chemistry-Geology pair. If he/she sends a query about mean of the pair, the other student's privacy can be breached.

E. Which cell should be suppressed to keep the privacy of students? State why.

주전공과 부전공 테이블

(Geology,Chemistry) 두명 노출되기 쉬우므로, suppression 시켜야 하며
같은 row에서 볼 때 (Physics,Chemistry), 그리고 (Physics,Biology),(Geology,Bilogy)를 suppression 시켜야만 추정하기가 어려워진다.

	Biology	Physics	Chemistry	Geology
Biology	-	16	17	11
Physics	7	-	32	18
Chemistry	33	41	-	2
Geology	9	13	6	-

	Biology	Physics	Chemistry	Geology
Biology	-	NA	17	NA
Physics	7	-	32	18
Chemistry	33	NA	-	NA
Geology	9	13	6	-

F. What is the limitation of the l-diversity? How this problem can be resolved?

Limitation of the L-diversity

[L-diversity 의 한계점]

1) Distinct l-diversity 인 경우 the probabilistic inference attacks 을 막지 못한다. 예를 들면, 하나의 동등 클래스에 열 개의 튜플이 있다. "Disease"속성 : 하나는 "Center" 다른 하나는 "Heart Disease", 나머지 8 가지는 "Flu"라고 가정하자. 이것은 3-diversity 를 만족하지만, 공격자는 공격목표가 되는 사람의 병이 "Flu"인지 확인할 확률이 80%가 된다.

2) l-diversity 는 속성을 쉽게 알 수도 있고, 어렵게 알 수도 있다.
예를 들면, 민감한 속성으로 HIV positive(1%)와 HIV negative(99%)의 값을 가진다고 가정할 때, 오진 HIV negative 값만 존재 하는 동등 클래스가 있다면 2-diversity 는 쉽게 알 수 있을것이다. 반대로, 1000000 개의 튜플이 있다고 쳤을 때 최대 100 개의 HIV negative 한 값이 존재한다면 분명한 2-diversity 를 가지게 되어 그 값을 쉽게 알 수 없을 것이다.

3) l-diversity 는 속성이 노출되는 것을 확실하게 막을 수 없다.

Similarity Attack

Bob	
Zip	*Age*
47678	27

A 3-diverse patient table

Zipcode	Age	Salary	Disease
476**	2*	20K	Gastric Ulcer
476**	2*	30K	Gastritis
476**	2*	40K	Stomach Cancer
4790*	≥40	50K	Gastritis
4790*	≥40	100K	Flu
4790*	≥40	70K	Bronchitis
476**	3*	60K	Bronchitis
476**	3*	80K	Pneumonia
476**	3*	90K	Stomach Cancer

[해결책]

1) t-closeness 방법을 사용한다.

l-diversity 기법에서는 속성값이 불균형하게 분포될 수 있는데 t-closeness 기법은 속성 도메인의 분류체계를 고려하여 고른 분포를 보장하였다. t-closeness 는 동등클래스의 민감한 속성의 분배를 전체 테이블의 민감한 속성의 분배와 가깝게 하는 것을 말한다.

3-diversity version

	ZIP Code	Age	Salary	Disease
1	476**	2*	3K	gastric ulcer
2	476**	2*	4K	gastritis
3	476**	2*	5K	stomach cancer
4	4790*	≥ 40	6K	gastritis
5	4790*	≥ 40	11K	flu
6	4790*	≥ 40	8K	bronchitis
7	476**	3*	7K	bronchitis
8	476**	3*	9K	pneumonia
9	476**	3*	10K	stomach cancer

linking attack (similarity attack)

0.278 (Disease)-closeness

	ZIP Code	Age	Salary	Disease
1	4767*	≤ 40	3K	gastric ulcer
3	4767*	≤ 40	5K	stomach cancer
8	4767*	≤ 40	9K	pneumonia
4	4790*	≥ 40	6K	gastritis
5	4790*	≥ 40	11K	flu
6	4790*	≥ 40	8K	bronchitis
2	4760*	≤ 40	4K	gastritis
7	4760*	≤ 40	7K	bronchitis
9	4760*	≤ 40	10K	stomach cancer

2) 치명적인 결합이 발생했을 때는 모조레코드를 삽입한다. 여기서 모조 레코드 개수를 저장하고 있는 보조 테이블도 따로 관리 해서 데이터 분석의 정확성을 높여야 한다.

[2010 중간고사 샘플 문제]

Which cell should be suppressed to keep the privacy of students? State why.

	Biology	Physics	Chemistry	Geology
Biology	-	16	17	11
Physics	7	-	32	18
Chemistry	33	41	-	2
Geology	9	13	6	-

Minimum query size = 4

Cell suppression 이란 특정한 Cell 이 query 의 결과에 나오지 않도록 차단하거나 암호화 시키는 방법이다. Minimum query size 4 보다 작은 2(Chemistry, Geology) 값을

suppression 시켜야 한다. 이 값을 유추하지 못하도록 41(Chemistry, Physics)를 suppression 하고, 11(Biology, Geology)를 suppression 한다.
또한 위의 suppression 한 값에 대한 유추를 하지 못하도록 16(Biology, Physics)로 suppression 해야 한다.

[2011 1학기 기말고사 기출]

아래 database 에 대해 4-anonymity 에 기반을 둔 clustering 을 적용할 때 혼합척도 X가 가능하면 가장 작게 되도록 이 database 를 분할하고 그때의 혼합척도 값을 쓰시오. (15 점)

Name	DOB	ZIP	Sex	Blood	Illness
Kennedy	1983	12335	F	A	Breast cancer
Einstein	1984	14567	M	B	HIV
Lee	1960	12346	M	O	Manic-depressive
Hanks	1963	51234	F	AB	Arteriosclerosis
Hanks	1965	51236	F	O	Angina pectoris
Wang	1968	51234	M	O	Colitis
Aitken	1976	51237	F	O	Mellitus
Taroni	1960	51236	M	A	Gastric hyperacidity
Tsomko	1954	14567	M	B	Empyema
Kim	1998	12347	F	O	Epilepsy
Yang	2000	51235	M	O	Hypotension
Tanaka	2001	14567	M	B	Precocious dementia

4. Lamport's time stamp, 프로세스와 clock time 문제

[2009년도 중간고사 기출]

- **Lamport's timestamps are used for synchronizing logical clocks in different processes. Explain how they work and give an example involving three processes that communicate with each other. What properties do the timestamps guarantee for the system?**

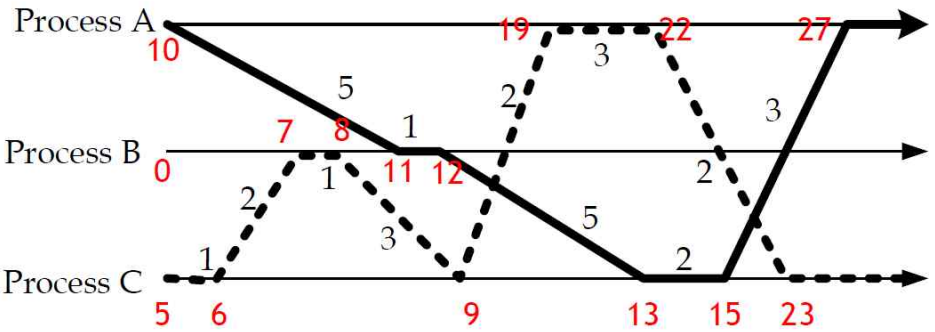
  Clock counter는 각 이벤트가 발생하기 전에 증가하고, 프로세스가 메시지를 보낼 때 메시지에 clock counter 값을 같이 보낸다. **메시지가 도착하면 받은 clock counter와 자신이 가지고 있던 clock counter와 비교하여 자기 것이 크다면 그대로 유지하고 새로 도착한 것이 더 크다면 새로운 clock counter에 1을 증가시킨 것을 자신의 clock counter로 설정**한다. 이러한 알고리즘을 전제로 event발생 주기보다 tick의 발생 주기가 더 빨라야 하며(하나의 프로세스에서 두 개의 사건의 동시 발생을 막기 위해), 또한 두 개의 프로세스가 동일한 시간에 이벤트 발생하는 것을 막기 위해 각 시간의 소수점 이하에 프로세스의 번호를 붙이는 형식을 가진다. 시스템이 이런 타임스템프를 보증되는 이유는 송신자가 수신보다 항상 빠르다는 특징이 있기 때문이다. Logical clock은 빠른지 느린지만을 판단할 뿐 실제로 얼마나 빠른지는 계산하지 않는다. 그렇기 때문에 이런 시스템적 특성만을 가지고 있어도 타임스템프를 충분히 보증할 수 있게 된다.

[2011봄 졸업 고사 기출]

- **All clock runs at the same rate but initially A,B, and C's clocks read 10, 0, and 5, respectively. Numbers in the figure show the time of transmission and time of processing.**

  A. Assuming that the figure in the next page implements Lamport timestamp system, assign the reasonable clock values and explain how the timestamps are obtained.

  > One solution is given as follows:

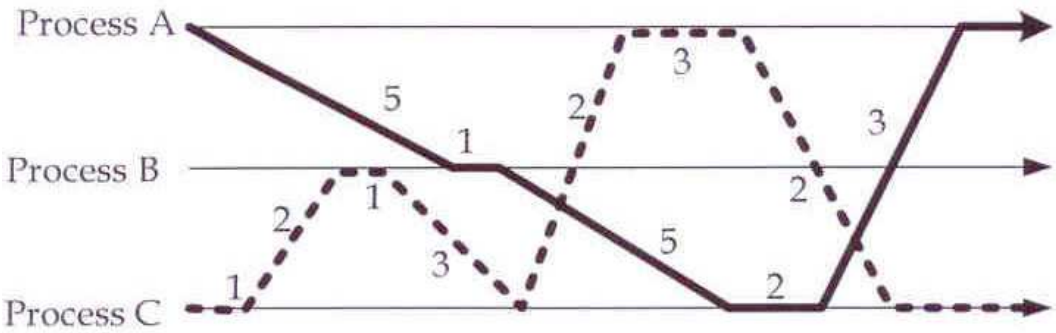


  under the assumption that all clock rates are synchronized and correct. Other solutions are available.

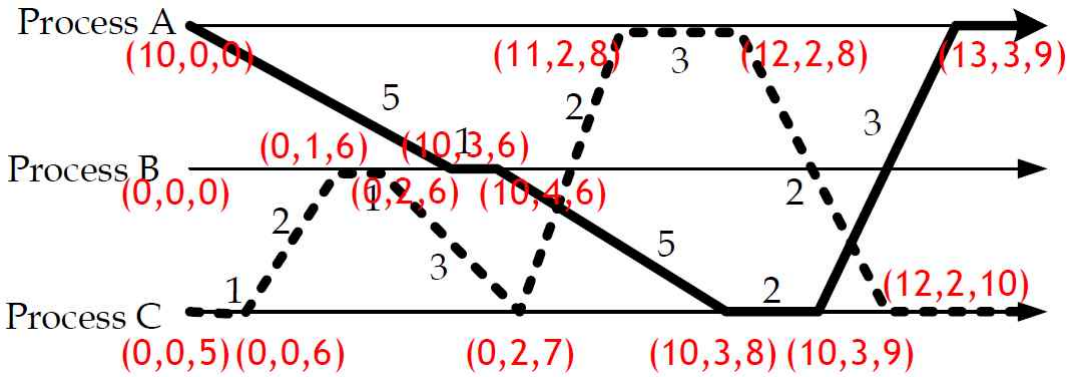
  B. What is the weakness of the Lamport logical system? Is the statement "if C(X) < C(Y), then X happened before Y" where C(X) is a clock of an event X? State why.

  > Lamport system does not work for the statement: C(A) < C(B), then A → B due to concurrent processes.

  C. Solve the weakness by using vector clocks.

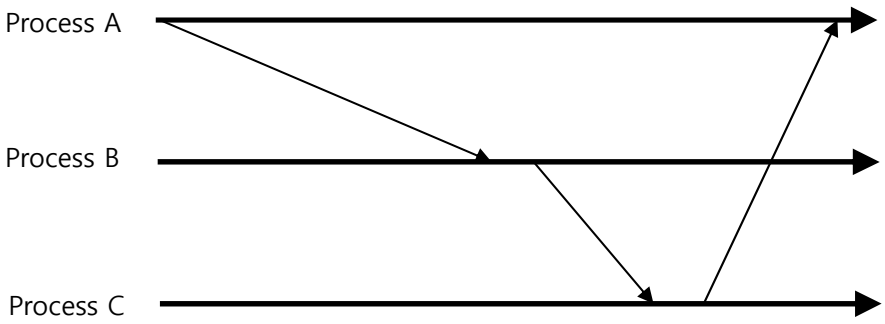


  > One solution is given as follows:



[2011 졸업시험 샘플문제]

Suppose there are three processes A, B and C. All clock runs at the same rate but initially A's clock reads 10, B's clock reads 0 and C's clock reads 5. At time 10 by A's clock, A sends a message to B, this message takes 4 units of time to reach B. B then waits one unit of time and then sends a message onto C which takes 2 units of time to reach C. C then waits one unit of time and then sends a message onto A which takes 2 units of time to reach A. Assuming that the system implements Lamport's timestamps, assign the reasonable clock values and explain how the timestamps are obtained in Figure 1. Arrows of thin lines indicate the transmission of a message, and thick lines the flow direction of time.



최종적으로 갖게 되는 clock value A=20, B=15, C=18

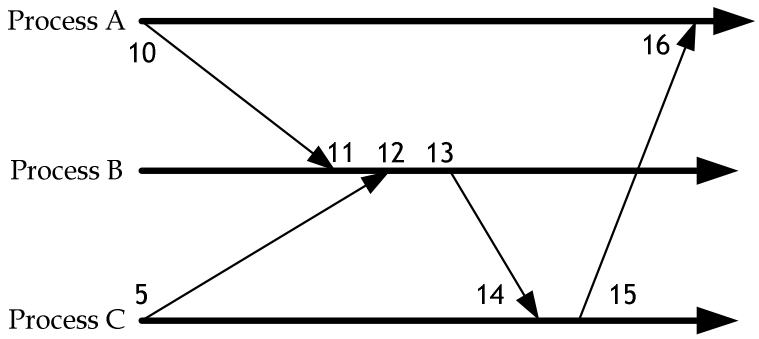
Ramport's timestamp는 최초의 process들이 고유의 clock value를 갖는 상태에서 어느 한 process가 다른 process에게 메시지를 보낼 때 자신의 clock value를 같이 보내고, 그것을 받은 process는 받은 clock value가 자신의 값 보다 큰 값이면 그것으로 교체하고, 자신의 것이 더 크면 자신의 값을 유지하는 방법이다. 위 그림에서 최초 메시지 전송으로 인해 B는 자신의 clock value 보다 큰 값인 A의 clock value에 전송시간과 지연한 시간을 포함하여 15로 clock value가 변경된다. C도 B로부터 받은 메시지의 timestamp value인 17이 자신의 clock value 5보다 크므로 timestamp 값을 변경하고 최종적으로 18을 갖게 된다.

[2010년 봄,가을 중간고사 기출]

- **Suppose there are three processes A, B and C. All clock runs at the same rate but initially A's clock reads 10, B's clock reads 0 and C's clock reads 5. At time 10 by A's clock, A sends a message to B, this message takes 4 units of time to reach B. B then waits one unit of time and then sends a message onto C which takes 2 units of time to reach C. C then waits one unit of time and then sends a message onto A which takes 2 units of time to reach A. Assuming that the system implements Lamport's timestamps, assign the reasonable clock values and explain how the timestamps are obtained in Figure 1. Arrows of thin lines indicate the transmission of a message, and thick lines the flow direction of time.**

A. Assuming that the figure in the next page implements Lamport timestamp system, assign the reasonable clock values and explain how the timestamps are obtained.

> See the following figure.



B. In the Lamport's clock mechanism, if A->B, then C(A) < C(B). However, its converse is not always true. State why the converse is not always true.

> A → B means A happens before B. C(A) < C(B) means that either A happens before B (A → B) or A and B are concurrent (A || B) where C(A) is smaller than C(B). Thus, if C(A) < C(B), it does not always mean A → B.

[과제] 문제를 정리해 보자면 3개의 프로세스 A,B,C가 있을 때, A는 10 B는 0 C는 5의 초기 값을 가진다. A가 10의 시간 위치에 있을 때 B로 메시지가 전해지고, 이는 4 개의 시간 유닛 만큼 소요된다. B는 메시지를 받은 후, 1개의 시간 유닛 만큼 기다렸다가 C로 메시지를 보내고, 이는 2개의 시간 유닛이 소모된다. 메시지를 받은 C는 1개의 시간 유닛만큼 기다렸다가 A에게 전송하고, 이는 2개의 시간 유닛이 걸린다. 이 내용을 표로 정리하면 다음과 같다

	Clock(A)	Clock(B)	Clock(C)
0	10	0	5
1	20	10	15
2	30	20	25
3	40	30	35
4	50	40	45
5	60	50	55
6	70	60	65
7	80	70	75
8	90	80	85
9	100	90	95
10	110	100	105

Event	TimeStamp	value
A	(1,0,0)	10
B	(1,1,0)	40
C	(1,2,0)	50
D	(1,2,1)	75
E	(1,2,2)	85
F	(2,2,2)	110

위의 표를 보면 time stamp 와 value 값 모두 적합한 것을 볼 수 있다. 하지만 이것은 항상 성립하는 것은 아니다.

하지만 At time 10 by 값을 2 로 주면 네 번째 에서 문제가 발생하는데 프로세스 A 번에서 14 에 전송된 메시지가 프로세스 B 번에 8 에 도착했다. 이전에 발생한 이벤트의 시간이 이후에 발생한 이벤트의 시간보다 크다.

	Clock(A)	Clock(B)	Clock(C)
0	10	0	5
1	12	2	7
2	14	4	9
3	16	6	11
4	18	8	13
5	20	10	15
6	22	12	17
7	24	14	19
8	26	16	21
9	28	18	23
10	30	20	25

Event	TimeStamp	value
A	(1,0,0)	10
B	(1,1,0)	8
C	(1,2,0)	10
D	(1,2,1)	19
E	(1,2,2)	21
F	(2,2,2)	30

위의 표에서 알 수 있듯이 시간 유닛의 값에 따라서 논리적으로 맞지 않는 상황이 벌어 질 수도 있다. 하지만 이러한 상황은 동기화(synchronization)과정을 통해서 해결할 수 있다. 메시지가 도착하면 받은 clock counter 와 자신이 가지고 있던 clock counter 를 비교하여 자신이 가진 것이

크다면 그대로 유지하고, 새로 도착한 것이 더 크다면 새로운 clock counter 에 1 을 증가시킨 것을 자신의 clock couter 로 셋팅 한다.

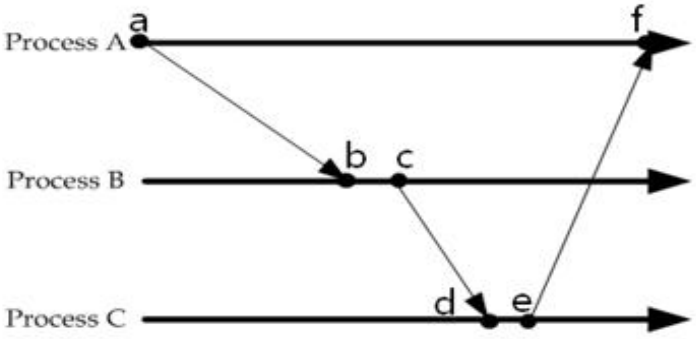
	Clock(A)	Clock(B)	Clock(C)
0	10	0	5
1	12	2	7
2	14	4	9
3	16	6	11
4	18	11	13
5	20	13	15
6	22	12	17
7	24	14	19
8	26	16	21
9	28	18	23
10	30	20	25

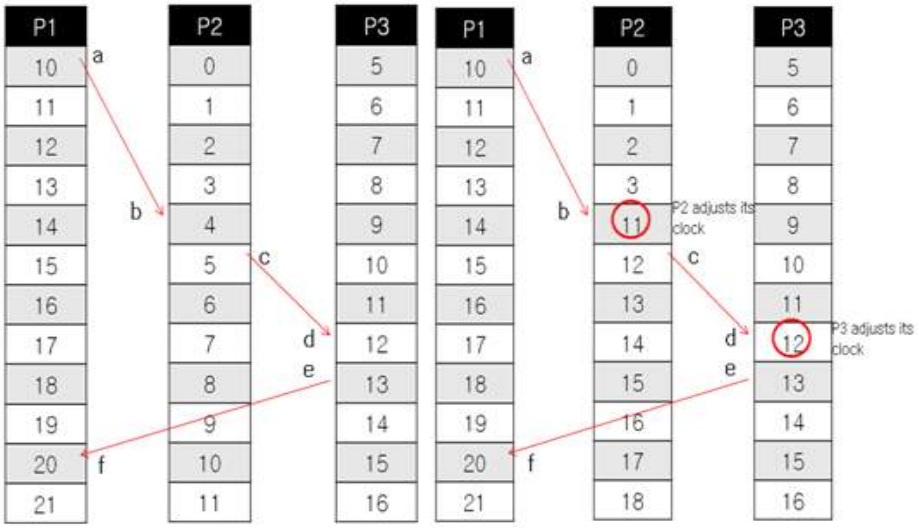
Event	TimeStamp	Value
A	(1,0,0)	10
B	(1,1,0)	11
C	(1,2,0)	13
D	(1,2,1)	19
E	(1,2,2)	21
F	(2,2,2)	30

Lamport Timestamp는 이벤트의 완벽한 동기화를 이루기 힘든 분산시스템에서 메시지 교환 중에 자신의 시스템의 카운터 값을 추가시켜 이를 통해 각 시스템이 동기화를 이루는 방법이다.

이 방법은 몇 가지 규칙을 따른다.

1. 각 프로세스는 이벤트 전에 카운터를 증가시킨다.
2. 프로세스가 메시지를 보낼 때, 메시지에 카운터 값을 포함시킨다.
3. 메시지를 받았을 때, 메시지의 카운터 값이 자신의 카운터 값 보다 크다면 자신의 카운터 값을 메시지 카운터 값 + 1로 세팅한다.





이제 위의 그림에서 timestamp를 설정해보자.

1. (a->b) P1은 자신의 카운터 값 10을 메시지와 함께 P2에 보낸다.

A. P2는 받은 값 10과 자신의 카운터 값 4를 비교하여 받은 값이 더 크기 때문에 받은 값에 1을 더하여 자신의 카운터 값을 11로 설정한다.

2. (c->d) P2는 1 unit of time만큼 기다린 다음에 P3에 메시지를 보낸다.

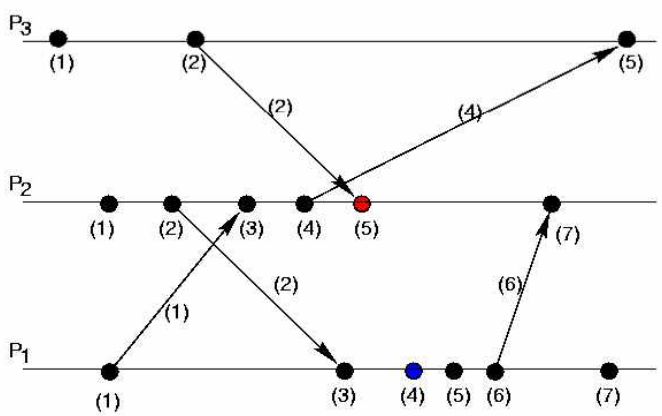
A. P3는 P2의 카운터 값이 자신의 값 보다 크지 않기 때문에 그대로 유지한다.

3. (e->f) P3는 P1에게 자신의 카운터 값 13과 메시지를 보낸다.

A. P1은 받은 값 13이 자신의 카운터 값 20보다 작기 때문에 자신의 값을 그대로 유지한다.

Show an example that if A B, then C(A) < C(B), but its converse is not true. How can you make the converse true?

1)



각각의 프로세스에서 이벤트 P3의 (2)에서 P2의 (5)로 향할 때 C(2) < C(5)임을 알 수 있다. 하지만 반대로 P2(5)를 P3(2)의 반대는 성립하지 않는 것을 알 수 있다.

2) Consideration > 하나의 프로세스에서 일어나는 이벤트를 각각 A, B라고 정의 했을 때, C(x)는 하나의 이벤트 'x'가 일어나는 타임스탬프(timestamp) 함수라고 하자. 함수를 정의하는 필요한 규칙들은 다음과 같다.

1. 논리적 시계(logical clock)가 적어도 하나가 존재하여 그 시계의 카운트 증가로 인한 차이로 이벤트 A와 B를 구분 한다.

2. 다중프로세스나 멀티쓰레드(multithreaded) 환경에서는 PID(process ID)나 다른 고유한 아이디를 타임스탬프에 주어 각각의 다른 프로세스에서 발생한 A와 B의 이벤트를 구분 한다.

Implications > 주어진 논리적 시계(logical clock)이 위의 규칙을 따를 때, 다음과 같은 관계가 성립된다.

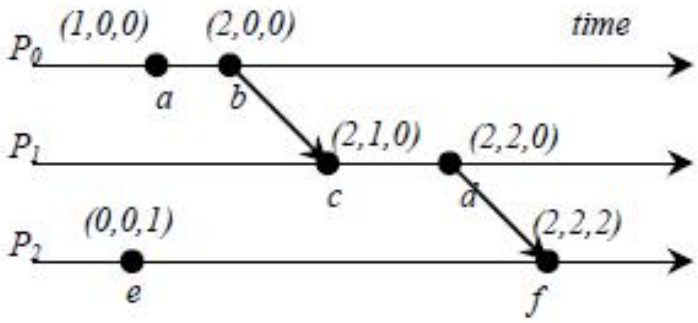
관계] 만약 Aà B 이면, C(A) < C(B) 이 성립한다. 여기서 ' à'의 의미는 "이전에 발생하였다(happened before)" 이다.

하지만 심플한 Lamport clock 쓰면, 부분적인 일반 순서만을 추측할 수 있다. 그 예로, C(A) à C(B)가 주어진다면 이벤트의 발생 순서가 A à B 임을 알 수 없다.

여기서, 대우(cotrapositive) 명제를 통해 C(A) (not)<C(B)의 의미는 $a \nrightarrow b$를 의미한다. 앞선 증명을 통해 $C(a) \geq C(b)$가 발생하게 되더라도, 이벤트 A는 B가 일어나기 전까지 발생할 수 없다. 더해서 C(A)<C(B)의 의미는 A의 이벤트는 B가 일어나는 시점과 같거나 전에 일어나는 것을 의미한다. 또한 B보다 일찍 발생할 수 없다.

Solution for this problem – Vector Clock

: 이 Clock 문제를 해결하기 위해 나온 것이 바로 Vector Clock이다.



이와 같이 프로세스마다 세 개의 Clock을 사용하여 나타내는 Vector Clock은 C(A) < C(B)이면, A -> B가 성립하는 것을 보장한다. e의 경우에는 f만 빼고 a, b, c, d와는 concurrent event이기 때문에 영향을 받지 않는다.

5. Lattice 와 Hasse diagrams문제

- **Lattice**

포셋의 임의의 두 원소 a, b에 대하여 한 개의 최소상한과 한 개의 최대 상한이 존재하면 그 포셋을 속(lattice)이라고 한다. Lattice 의 두 가지 조건은 임의의 두 원소 a,b에 대하여 한 개의 최소상한과 한 개의 최대상한이 존재한다.

1) Existance of binary joins
   - Least upper bound(joins)가 존재한다.

2) Existence of binary meets
   - Greatest lower bound(meets)가 존재한다.

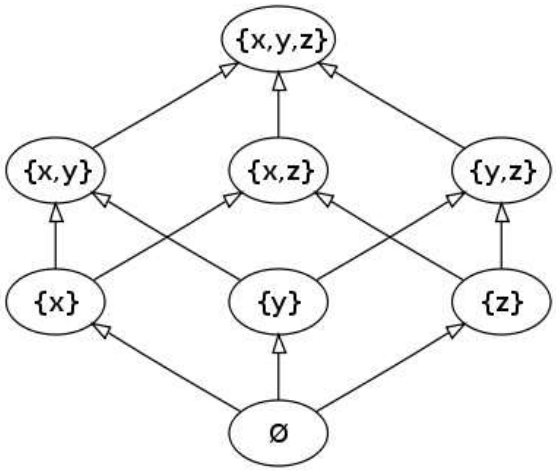
최소 상한계(Least upper bound)
   - 포셋 내 부분집합 A의 다른 모든 상한보다 작은 상한이면 A의 최소상한

최대 상한계(greatest lower bound)
   - 포셋 내 부분집합 A의 다른 모든 하한보다 큰 상한이면 A의 최대하한

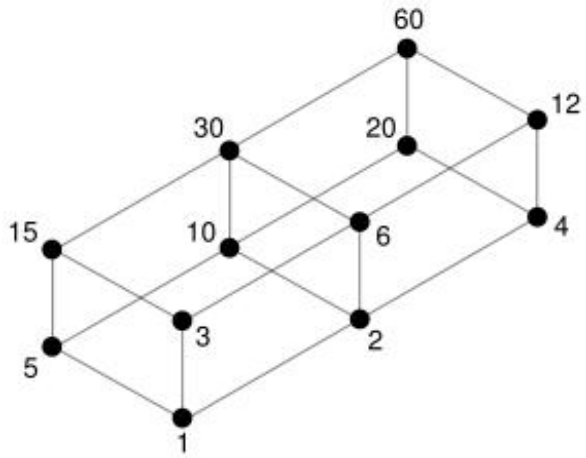
- **Hasse diagram**

히세 도표는 부분순서집합의 원소를 표현하기 위해 고안된 표기법으로, 각 원소의 순서 관계를 그래프로 표현한다. 하스 그림의 노드들은 포셋 A의 요소들을 나타낸다. X < Y 라면 노드 X는 노드 Y밑에 놓고 X와 Y는 한선(edge)에 의해 연결된다. 하스그림은 본래의 부분적 순서에 대한 많은 정보를 포함한다. 즉, 어느 요소가 다른 요소의 선행 요소라는 사실을 그 요소들이 연속된 선들로 알 수 있다.

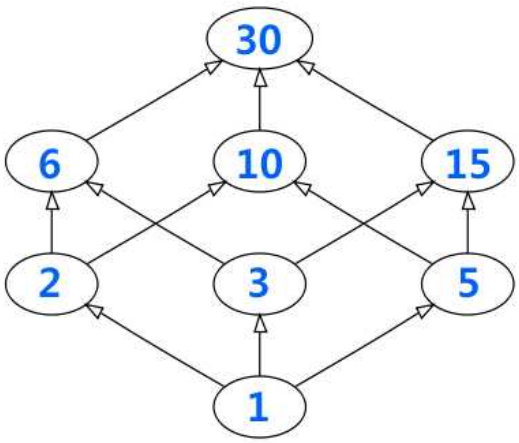
Ex) 집합 {a,b,c}의 멱집합 원소들에 대해서 부분집합 관계는 다음과 같이 표현된다.



Ex) 60의 약수에 대해서 배수 관계는 다음과 같이 표현된다. The set A={1,2,3,4,5,6,10,12,15,20,30,60} of all divisors of 60, partially ordered by divisibility, has the Hasse diagram :



Ex) Let D_{30} denote the set of all positive divisors of 30. Then D_{30}={1,2,3,5,6,10,15,30}. Show that this poset is a lattice.



D_{30} 을 Hasse Diagram으로 표현하면 위와 같은 그림이 나오게 된다. 그럼 위의 조건을 통하여

Lattice 인지 아닌지 판단해보자.

*** MINIMAL ELEMENT**

b ∈ D*30*
b ≠ 1, b < 1, b ∤ 1 를 만족하는 원소는 없다.

따라서 1은 minimal element 이다.

*** LEAST ELEMENT**

b ∈ D*30*
1 ≤ b, for all b ∈ D*30*를 만족하기 때문에 (1은 모든 원소를 나눌 수 있다.)

1은 least element 이다.
(2는 3,5,15를 / 3은 2,5,10을 / 5는 2,3,6을 나누지 못한다. 따라서 least element 가 아니다.)

*** MAXIMAL ELEMENT**

b ∈ D*30*
b ≠ 30, b > 30, 30 ∤ b 를 만족하는 원소는 없다.

따라서 30은 Maximal element 이다.
(6, 10, 15는 30을 나누기 때문에 maximal element가 아니다. 2,3,5 또한 그렇다.)

*** GREATEST ELEMENT**

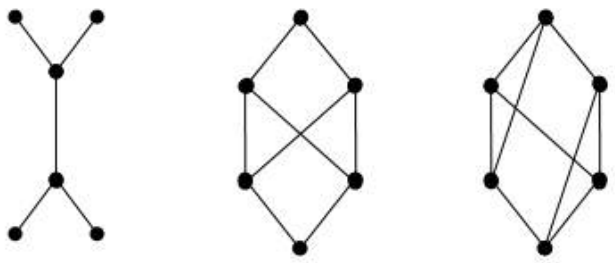
b ∈ D*30*
b ≤ 30, for all b ∈ D*30*을 만족하기 때문에 (모든 원소는 30을 나눌 수 있다.)

30은 Greatest element 이다.

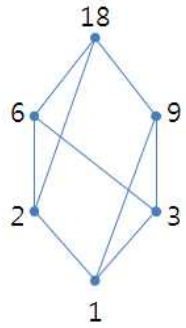
1 은 greatest lower bound 이며, 30은 least upper bound로서 Lattice의 성질을 만족한다.

[2010가을 기말고사 기출]

- **Decide which of the following graphs are lattice.**



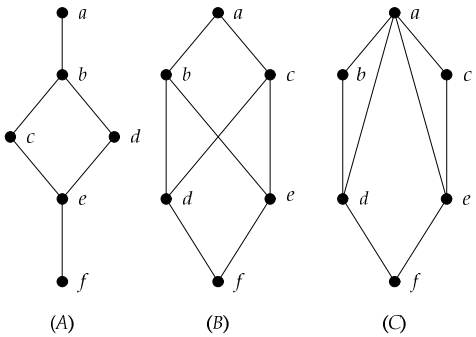
첫 번째 도형은 lattice가 성립하지 않는다. 두 번째 도형 역시 성립하지 않는다. 세 번째 도형을 lattice 성립한다. D_{18}을 넣어보면D_{18} = { 1,2,3,6,9,18 }



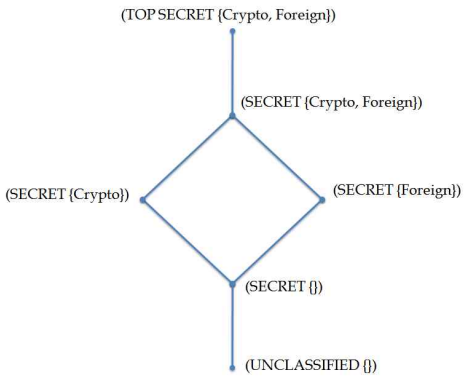
1 은 greatest lower bound 이며, 18은 least upper bound로서 Lattice의 성질을 만족한다.

[2011 봄 종합시험 샘플문제]

- **Determine whether the posets represented by each of the Hasse diagrams below are lattices. If it is not a lattice, describe why not.**

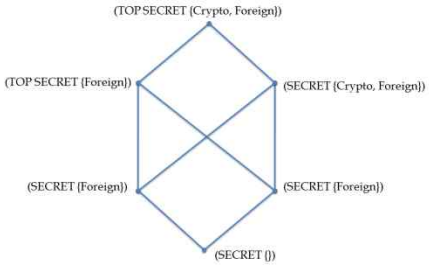


(a)



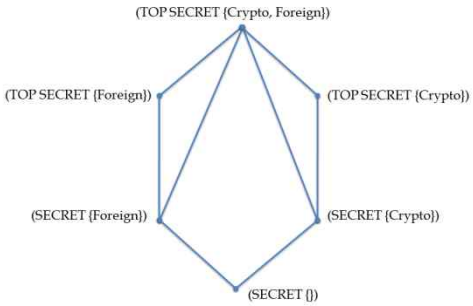
SECRET Level은 Unclassified, Secret, Top Secret으로 나뉜다. 우선 최 하단에 Unclassified가 위치되고 이 위치를 기준으로 상단에 Secret레벨이 위치하게 된다. 그 다음은 2갈래로 갈라지게 되므로, Secret의 메서드인 Crypto와 Foreign을 위치시킨다. 그 위 취합되는 부분은 Secret이 두 메서드를 가지도록 구성한다. 마지막으로 최 상단에 TopSecret을 위치한다. 따라서 Lattice 만족한다.

(b)



Secret{foreign}이 동일한 레벨에서 Conflict 된다. 따라서 Lattice 아님

(c)



Secret으로 시작하여 좌우는 Foreign, Crypto메서드를 가진 Secret으로 나뉘게 되며 그 위에 Top Secret또한 각각 Foreign, Crypto를 가지게 된다. 최 상위 노드는 이 두 메서드를 포함하게 되며 TopSecret레벨이 된다. 따라서 Lattice 만족한다.

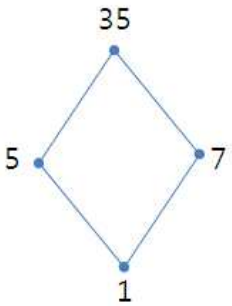
[2010 기말 샘플 문제]

- Let D_{35} denote the set of all positive divisors of 35.

  A. What are elements of this set?

  D_{35} = {1, 5, 7, 35}

  B. Draw a Hasse diagram.

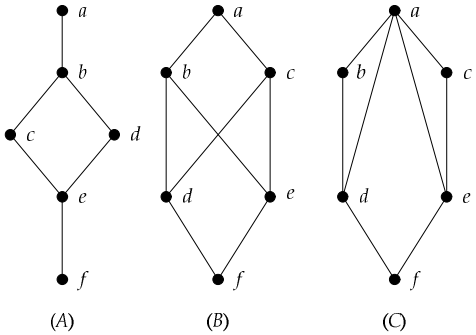


C. Verify whether it is a lattice or not.

1 은 greatest lower bound 이며, 35은 least upper bound로서 Lattice의 성질을 만족한다.

[2009가을 기말 샘플 문제]

- **Determine whether the *posets* represented by each of the *Hasse diagrams* below are lattices. If it is not a lattice, describe why not.**



Sol) Lattice는 poset에서 유일한 Supermum(Least Upper bound)과 Infimum(Greatest Lower bound)이 존재하는 것을 뜻한다. (A)는 Poset을 도식화한 Hasse diagram으로서 Supermum(a)와 Infimum(f)가 존재한다. **따라서 (A)는 Lattice이다.** (B)는 Poset을 도식화한 Hasse diagram이지만 b와 c는 유일한 Infimum이 존재하지 않는다. 왜냐하면, d,e,f 각각은 b와 c의 Lower bound이지만 이 Poset의 ordering에 대하여 d,e,f중 어떤 원소도 다른 두 원소를 앞서지 못한다. 따라서, b와 c는 유일한 Infimum을 갖지 않는다. 같은 이유로 d와 e 또한 유일한 Supermum이 존재하지 않는다. 따라서, **(B)는 Lattice가 아니다.** (C)는 Supermum(a)와 Infimum(f)가 존재하지만, Hasse diagram은 reflexivity property에 의한 loop와 transitivity property에 의한 a <= b, b<=c => a<=c 일때, a와 c를 연결하는 선은 긋지 않는다. 따라서, (a,d), (a,e)를 잇는 선이 존재하므로 (C)는 Hasse diagram이 아니다. 또한 b와 c는 (B)와 같은 이유로 유일한 Infimum 이 존재하지 않는다. 따라서, **(C)는 Lattice가 아니다.**

6. *-property, bibamodel, bel 라파듈라 모델

[2010가을 기말고사 기출]

- **Describe about what ss- and *-property of Bell-LaPadula, Biba, and Chinese wall models are.**

*-property 이란

bell-lapadula 모델에서는 NWD – No Write down 고 등급 주체가 저 등급 기밀성 객체에 정보를 쓰는 것을 불허 하는 것을 뜻한다.

접근제어 모델은 크게 임의적 접근제어모델(DAC)과 강제적 접근제어모델(MAC)로 분류되며
DAC는 정보의 소유자가 정보의 보안레벨을 결정하고 이에 대한 정보의 접근제어도 설정하는 모델이며, DAC은 무척 편리한 접근제어모델이지만 파일의 소유자, 즉 정보의 소유자가 정보의 보안레벨과 접근제어를 설정하기 때문에 중앙 집중화된 정보의 관리가 어렵다. 따라서 정보에 대한

엄격한 접근제어는 사실상 거의 불가능하다. 강제적 접근제어 모델(MAC)은 중앙에서 정보를 수집하고 분류하며, 각각의 보안레벨을 붙이고, 이에 대해 정책적으로 접근제어를 수행한다.

✓ **Bell-LaPadula model(BLP)**

최초의 수학적 모델로 알려져 있고, 군대의 보안레벨과 같이 그 정보의 기밀성에 따라 상하 관계가 구분된 정보를 보호하기 위해 사용한다.

읽기 권한과 쓰기 권한

- 읽기 권한 : 낮은 보안 레벨의 권한을 가진이가 높은 보안 레벨의 문서를 읽을 수 없으나, 자신의 권한보다 낮은 수준의 문서는 읽을 수 있음. ➔ NRU ss property : simple property security

- **자신보다 높은 보안 레벨의 문서에 쓰기는 가능하지만 보안레벨이 낮은 문서에는 쓰기 권한이 없음 (이를 * property라고 함) ➔ NWD *property : start(*) property security**

✓ **Biba model**

기밀성보다는 좀 더 신뢰할 수 있는 정보, 즉 정보의 무결성을 높이는 데 목적이 있는 경우에 사용

읽기 권한과 쓰기 권한

읽기 권한: 무결성 레벨 2인 사람이 무결성 레벨 1을 읽을 수 있고, 무결성 레벨 3인 정보는 읽을 수 없음.

쓰기 권한: 무결성 레벨 2의 문서를 더 높은 신뢰도를 가진 무결성 레벨 1의 정보에 쓸 수 없음. 읽기와 마찬가지로 낮은 신뢰도를 가진 정보를 높은 신뢰도를 가진 정보에 더해, 결과적으로 높은 신뢰도를 가진 정보의 신뢰도를 떨어뜨리지 않는 것임. 하지만 높은 신뢰도를 가진 무결성 레벨 2의 정보를 신뢰도가 더 낮은 무결성 레벨 3의 문서에 쓰는 것은 가능.

: Integrity(무결성)에 관련된 간단한 모델이다. Biba model은 두 가지 보안 규칙이 있는데

1) The simple integrity axiom(단순 무결성 공리)

: 주체 S는 자신보다 낮은 무결성 수준을 가진 객체 O를 읽을 수 없다.

➔ NRD : No read down = simple property security

2) The * integrity axiom(스타 무결성 공리)

: 주체 S는 자신보다 높은 무결성 수준을 가진 객체 O에 쓰기를 할 수 없다.

➔ NWU : No write up

Biba 모델에서 제안한 보안 규칙의 의미를 알아보면,

1)번 성질은 여기서 객체의 무결성이 낮다는 것은 그 내용이 바뀔 가능성이 높도록 유지되고 있다는 말이다. 예로 아무나 쓸 수 있는 게시판의 글은 다른 누군가가 그 내용을 바꿀 수도 있다. 반면에 정식 공문은 여러 단계에서 결재를 거치게 되므로 제 3 자가 쉽게 바꿀 수 없다. 즉 무결성이 낮은 객체는 신뢰도가 떨어지는 정보를 가지고 있다는 의미이고 잘못된 정보를 보유할 가능성이 높다는 말이 된다. 따라서 자신보다 무결성 수준이 낮은 객체로부터 정보를 읽게 되면 잘못된 정보를 얻을 가능성이 생긴다.

2)**번 성질**은 자신보다 무결성 수준이 높은 객체로 정보를 기록하게 되면 그 객체의 무결성이 떨어 질 수 있다. 예로 그 객체는 원래 160비트 해시를 사용하여 무결성이 제공되고 있는데 취급자는 128비트 해시를 사용할 수 밖에 없다고 한다면 바뀐 정보는 128비트 해시를 가지게 되므로 변경된 내용은 이전보다 낮은 무결성을 가지게 된다.

chinese wall model(chapter 9)

① 개념

- to avoid conflicts of interest -> 충돌을 야기시키는 어떠한 정보의 흐름도 없어야 한다
- 동일한 영역에 있는 다른 회사의 자료에 접근해서는 안됨(직무분리)

② 적용 영역

- 투자, 금융, 광고 분야의 컨설팅

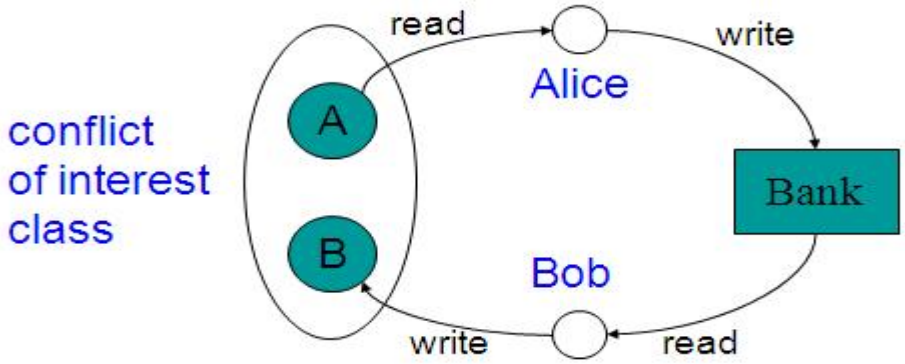
(ex) 컨설턴트 자신이 컨설팅하는 업체의 정보만 접근

③ 적용 예

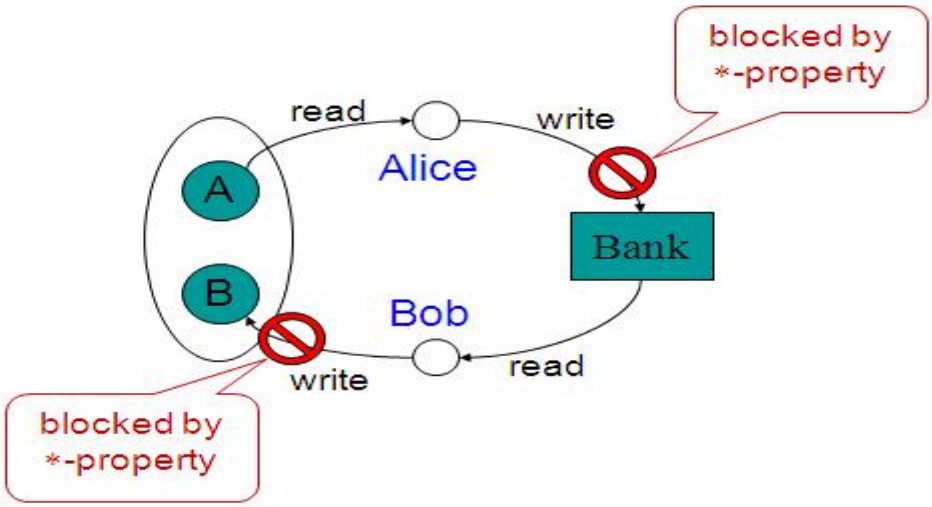
간접적인 정보흐름(A와 B는 같은 하나은행에 계좌를 가진 경쟁사이다.)

A와 하나은행에 거래하는 Alice는 A에 관한 민감정보를 하나은행 포트폴리오에 업데이트하려한다.

B와 하나은행에 거래하는 Bob은 하나은행 포트폴리오에 있는 경쟁자에 대한 정보에 접근하려고 한다.



이와 같은 경우에 chinese wall property에 의해 Alice는 A에서 읽어온 정보를 은행 포트폴리오에 업데이트할 수 없고, Bob의 경우 읽어온 정보를 B에게 쓰기할 수 없다.



[접근 통제 보안 모델 한번에 요약]

- 비공식적인 모델 ------- 수학적 검증 -----→ 공식적 보안 모델
  명백, 실행 가능, 쉽게 이해, 정책 반영

(1) 벨 라파듈라 모델(BLP)

- 군사용 보안구조의 요구사항을 충족하기 위해 설계된 모델
- 가용성이나 무결성 보다 비밀유출(Disclosure, 기밀성)방지에 중점
- MAC기법
- 최초의 수학적 모델

① 속성

- No Read Up(NRU or ss-property)
  : 보안 수준이 낮은 주체는 보안수준이 높은 객체를 읽어서는 안됨

② 제한 사항

- 접근권한 수정에 관한 정책이 없다.
- 은닉채널을 포함
- 기밀성만 다루고 있고 무결성은 다루지 않는다.

(2) 비바모델(BIBA)

- 무결성을 강조한 최초의 수학적 모델
- BLP모델의 단점인 무결성을 보장할 수 있도록 보완한 모델

  * DAC는 MAC의 단점을 보완한 기술이 아니다.

① 속성

- No Read Down(NRD or Simple Integrity Axiom)

  : 보안 수준이 높은 주체는 보안 수준이 낮은 객체를 읽을 수 없다.
- No Write Up(NWU or *Integrity Axiom)

  : 보안 수준이 낮은 주체는 보안수준이 높은 객체에 기록해서는 안됨

(3) 클락/윌슨 모델(Clark & Wilson model)

① 정의

- Well-formed transactions, separation of duties(직무 분리)

② Addresses all 3 integrity goals

- 비 인가자가 수정하면 안됨
- 직무분리 : 내부 일관성, 외부 일관성
- 권한 있는 사람이 부적절한 수정을 하면 안됨

③ Consists of Triples(Triple Access) : Subject/Program(신뢰성,일관성)/Object)and Rules

④ 예측가능 하고 완전한 방식으로 일어나야 함(Well-Formed Transaction)

(4) 만리장성

① 개념

- to avoid conflicts of interest -> 충돌을 야기시키는 어떠한 정보의 흐름도 없어야 한다.
- 동일한 영역에 있는 다른 회사의 자료에 접근해서는 안됨(직무 분리)

② 적용 영역

- 투자, 금융, 광고, 분야의Consulting / 컨설턴트 자신이 컨설팅 하는 업체의 정보만 접근

7. 생일 역설 문제 (birthday pradox)

- **What is the birthday paradox? What is the implication of this paradox? What is the minimum number of people who need to be in the room so that the probability that at**

least two of them have the same birthday is greater than $\frac{1}{5}$?

생일 문제란 확률론에서 유명한 문제로, 몇 명 이상 모이면 그 중에 생일이 같은 사람이 둘 이상 있을 확률이 충분히 높아지는지를 묻는 문제이다. 얼핏 생각하기에는 생일이 365~366가지이므로 임의의 두 사람의 생일이 같을 확률은 1/365~1/366이고, 따라서 365명쯤은 모여야 생일이 같은 사람이 있을 것이라고 생각하기 쉽다. 그러나 실제로는 23명만 모여도 생일이 같은 두 사람이 있을 확률이 50%를 넘고, 57명이 모이면 99%를 넘어간다. 이 사실은 일반인의 직관과 배치되기 때문에 생일 역설이나 생일 패러독스라고도 한다.

$$\begin{aligned}\bar{p}(n) &= 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{n-1}{365}\right) \\ &= \frac{365 \times 364 \cdots (365 - n + 1)}{365^n} \\ &= \frac{365!}{365^n (365 - n)!}\end{aligned}$$

가 되고, 최종적으로 구하고자하는 생일이 같은 사람이 둘 이상 있을 확률 p(n)은

$$p(n) = 1 - \frac{365!}{365^n (365 - n)!}$$

가 된다. 여기서, n≤365 인 자연수이고, !는 계승을 의미한다. 이 p(n)값을 특정 n 값에 대해 계산하면, 다음과 같다.

n p(n)
10 12%
20 41%
30 70%
50 97%
100 99.99996%

n=20일 때 41%의 확률, n=30일 때, 70%, 그리고 n=50이면, 97%로 계산된다. 즉, 50명만 모이면 그 가운데 2명 이상의 생일이 같을 확률이 97%이고, 100명이 모이면 거의 1에 가까워진다는 것을 알 수 있다.

[문제해결]

$$p(n) = 1 - \frac{365!}{365^n (365 - n)!}$$

$$1 - \frac{365!}{365^n(365-n)!} = 0.2$$

$$\frac{4}{5} \geq \frac{365!}{365^n(365-n)!}$$

$$4.365^n(365\text{-}n)! \geq 5.365!$$

$$4.365^n \geq \frac{365!}{365^n(365-n)!}$$

계산을 해보면, n=13명일 때 19% n=14명일 때 22%나온다 따라서,

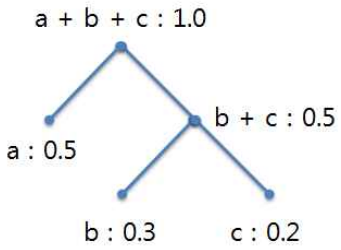
14명일 때 $\frac{1}{5}$ 이상 나온다.

8. 허프만 코드

- **Given probability of symbols a, b and c with 0.40, 0.30, and 0.30, respectively, encode the sequence $ababc$ using Huffman coding and arithmetic coding in binary format.**

Huffman coding :

a, b, c를 허프만 트리로 묶으면 다음과 같다.



각 문자에 대해 허프만 코드표를 작성하면 다음과 같다.

Data	Code
A	0
B	10
C	11

***ababc*를 허프만 코드로 변환하면 다음과 같다.**

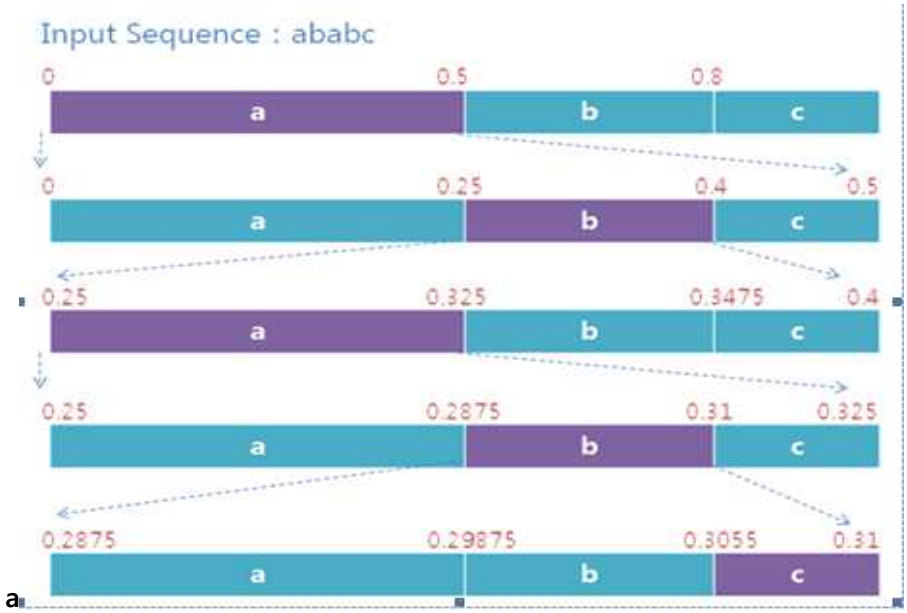
ababc = 01001011

Arithmetic code :

a, b, c 각각에 대한 확률은 다음 표와 같다.

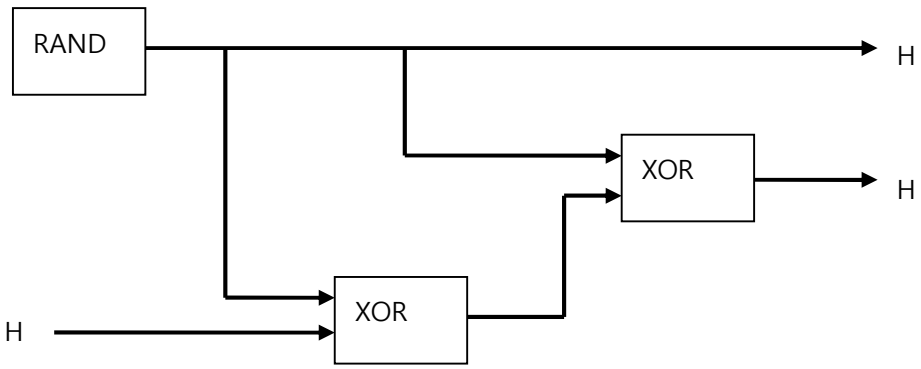
Data	Prob.
A	0.5
B	0.3
C	0.2

***ababc*를 Arithmetic code로 변환하면 다음과 같다.**



Termination : Encode the lower end (0.3055) to signal the end.

9. 기타 문제



위의 시스템을 식으로 풀어보면

(H1 ⓧ H2) ⓧ H2 = H3 인데, 좌변의 식을 풀어보면,

(H1 ⓧ H2) ⓧ H2 = H1

∴ H1 = H3

즉, 입력한 H1이 그대로 출력으로 나오게 되므로 위의 시스템은 안전하지 않다.

- **RSA algorithm is summarized as follow.**
  1) **Choose two large primes, p and q.**
  2) **Compute $n = p \times q$ and $z = (p-1) \times (q-1)$.**
  3) **Choose a number relatively prime to z and call it d.**
  4) **Find e such that $e \times d = 1 \bmod z$.**

  For $p = 7$ and $q = 11$, find suitable values d and e.

풀이>

p = 7, q = 11

n = 7 × 11 = 77, z = (7 − 1) × (11 − 1) = 60

choose number **d = 13**, d is gcd(60,d)=1 ,

then 13 × $e = 1 \bmod 60$

q	r_1	r_2	r	t_1	t_2	t
4	**60**	**13**	8	0	1	-4
1	13	8	5	1	-4	5
1	8	5	3	-4	5	-9
1	5	3	2	5	-9	14
1	3	2	1	-9	14	-23
2	2	1	0	14	-23	60
	1	0		**-23**	60	

-23 ≡ 37 mod 60

∴ **Choose e = 37**

Answer : d = 13, e = 37

- **Explain Needham-Schroeder protocol and answer if this protocol secure against replay attack? State why.**

1) 프로토콜 정의

. S : 인증서버 . A : Alice . B : Bob . KA : Alice 비밀키(Encryption Key)

. KB : Bob의 비밀키(Encryption Key) . Ks : 세션키 . RA, RB : 랜덤넘버

2) 프로토콜 분석

(1) Alice --> S

엘리스가 밥과 통신하기를 원하여 A,B,RA를 보낸다

(2) S -->Alice :EA(RA, B, K, EB(K,A))

S는 랜덤 세션키를 생성하고 그 세션키와 Alice와 Bob의 비밀키로 각각을 암호화하여 엘리스에게 보낸다.

(3) EB(K, A)

Alice는 메세지와 세션키 k를 구하고 S에게 보낸 RA를 검증 후 S가 Bob의 비밀키로 암호한 메세지를 보낸다.

(4) EK(RB)

Bob은 비밀키로 메세지와 세션키 K를 구하고 또다른 랜덤넘버 RB를 생성하여 세션키 K로 암호화하여 Alice에게 보낸다.

(5) EK(RB-1)

Alice는 세션키 K로 메세지를 구하고 RB-1을 랜덤넘버를 생성하여 세션키 K로 암호화하고 Bob에게 보낸다.

(6) Bob은 세션키 k로 메세지를 구하고 랜덤넘버 RB-1을 검증한다.

* 여기서 랜덤넘버 RA, RB, RB-1은 Replay attack을 방지하기 위한 것이다.

(7) 그럼에도 불구하고 세션키의 타임에 대한 취약점과 (2) 과정에서 공격자가 세션키를 가로챈다면 Bob은 공격에 당할 수 밖에 없다.

[용어 정리]

[2011봄 졸업고사 기출]

- **Regarding social networking, answer the following questions. (20points)**
  - **A. Compare click fraud and impression spam.**
    - ✓ **클릭사기(click fraud)** - 사용자 클릭 수를 증가시키기 위해 웹 사이트에 광고를 할 때 발생하는 부정 행위. 광고 사이트에 수동으로 연결하게 하거나 소프트웨어나 자동 프로그램을 이용하여 광고 클릭을 하도록 하는 방법을 쓴다. 클릭 사기는 주로 개인의 광고 수입을 올리기 위해 개인이 범하는 것과 경쟁사의 광고 비용을 소모시키기는 수단으로 사용하기 위해 기업이 범하는 것이 있다.
      웹사이트의 광고의 click 수를 증가시키는 software를 이용하여 클릭 수를 불법으로 발생시켜 광고주의 광고비를 고갈되게 하는 불법적인 행위이다.
    - ✓ **노출스팸(Impression spam)** – 자신과 동일한 keyword로 경쟁하는 업체가 있을 경우, 자신의 광고를 내린 후에 해당 keyword를 반복적으로 검색에 이용되도록 하여, 경쟁자가 검색에는 노출되지만 click으로 이어지지 않아 광고 순위가 낮아지게 되어, 자신

의 사이트를 상위에 rank 시키는 것이다.

B. What are Google bomb and Google hack?

- ✓ **구글폭탄(구글폭격)** – 검색에서 상위노출시키는 다양한 Black Hat기법을 사용하여 특정사이트를 상위에 노출시키지만 막상 해당사이트에는 검색한 키워드와는 상반되거나 전혀 다른 사이트로 검색자 들을 유인하는 것을 말한다. 또 다른 말로 검색사이트 구글에서 특정 웹사이트의 페이지순위 점수를 높여 검색결과가 상위에 오르도록 조작하는 방법을 가리킨다. 검색엔진의 알고리즘이 가진 맹점을 악용하는 기술이다. 이 기술을 이용해 구글의 검색결과를 조작하는 행위를 구글폭격(Google Bombing)이라고 한다.
- ✓ **Google hack -** The bad guys use search engines to find web servers that are running vulnerable applications (see Google hacking). 구글 검색과 구글의 어플을 이용해서 컴퓨터 코드나 웹사이트의 보안상 취약점을 찾아내는 방법을 말한다.

C. What are opt-in and opt-out?

- ✓ **Opt-in** – 광고성 이메일을 받기로 사전에 선택한 것
- ✓ **Opt-out** – e-mal을 보내서 받은 사람이 수신을 거부하면 이후에는 보낼 수 없도록 하는 것

[2009 2학기샘플 기출]

- **What are four things to come together for good security engineering framework? Briefly explain those four things.**

1) Integrity(무결성) - 데이터 및 네트워크 보안에 있어서 정보가 인가된 사람에 의해서 만이 변경 가능하다는 확실성이다. 정보는 노출되어 있지만 정보(데이터)의 변조를 방지하는 기능이다. 전자결제 같은 경우 금융결제원의 공인인증서를 통해 무결성을 보장 받는다.

2) Confidentiality(기밀성) - 정당한(합법적인) 사용자가 아닌 사용자들은 컴퓨터 시스템상의 데이터 또는 컴퓨터 시스템 간에 통신 회선을 통하여 교환, 전송되는 데이터의 내용을 볼 수 없게 하는 기능이다. 쉽게 말해 정보의 노출 방지를 위해 허가된 사용자만 접근 할 수 있도록 하는 것이다.

3) Availability(가용성) - 정보의 필요시 사용할 수 있게 하는 기능이다.

4) Non-repudiation(부인방지) - 메시지의 송수신이나 교환 후, 또는 통신이나 처리가 실행된 전송된 메시지에 대한 분쟁을 방지해준다. 예를 들어 인터넷뱅킹을 통해 예금을 인출 한 후 인출한 사실에 대해 부인을 할 경우를 막는 것이다. 인터넷 뱅킹의 경우 디지털서명을 통해 부인방지 기능을 수행하고 있다.

- **What is the buffer overflow attack? What would be a solution to prevent this kind of**

attack? What is the role of canary?

Buffer overflow attack은 지정된 크기의 저장 공간(buffer)보다 넘치게(overflow)입력되는 overflow 현상을 이용하여 buffer에 입력되지 못한 조작된 데이터를 시스템의 특정 위치에 기록하여 공격하는 기법이다. Buffer overflow attack을 막기 위한 방법으로 크게 두 가지가 있다. Stack을 non-executable하게 만드는 방법과 Stack의 레이아웃을 조정하는 방법이 있다. 이때, 두 번째 방법은 Stack guard와 Stack shield가 있다. stack guard는 canary를 스택에 넣어서 overflow를 알아내는 방법이고, stack shield는 리턴주소(return address)를 별도의 자료 구조에서 관리하는 방법이다. Stack 기반 overflow 기법이 지역 변수 근처의 return address를 변조한다는 것을 알고 return address 근처에 Canary라 불리 우는 4Byte 변수를 삽입하는 것이다. Canary는 함수의 역할이 끝난 후 이전 함수로 돌아갈 때, Canary 변수의 변경 여부를 검사하여 공격이 발생하였는지를 판단하는 것이다.

[2010 2학기기말 기출]

- **Describe about what ss- and *-property of Bell-LaPadula, Biba, and Chinese wall models are. (10pt)**
- **What is a block cipher and what is a stream cipher? State pros and cons of these methods. Why the stream ciphers are preferred in mobile and wireless environment?**

  Block Cipher : 평문을 정해진 size로 나누어진 Block 별로 암호화 하는 방법이다.

  장점 : 여러가지 block mode의 사용으로 안정성이 높고,(복잡한 구조) stream cipher의 key stream 생성기로도 사용이 가능하다.

  단점 : block 이 다 도착한 후에 암호화하므로 느리고 실시간 시스템에 적합하지 않다.

  Stream cipher : 평문을 작은 크기로 나누어, key generator에서 생성되는 key stream 으로 암호화 하는 암호화 방식이다.

  장점 : data가 들어오는 대로 암호화 하면 되기 때문에 빠르다.

  단점 : 작은 크기 단위로 암호화를 하기 때문에 암호화 횟수가 많아져 비효율적이다.

  Mobile & wireless : mobile & wireless 환경에서는 data의 오류가 자주 나타나게 되는데 이 때문에 error propagation이 되지 않는, stream cipher를 사용하는 것이 유리하고, byte 단위로 빠르게 암호화가 되기 때문에 stream cipher가 적합하다.

- **State what Kerckhoff's principle is. Explain briefly why a cryptosystem designed by someone who follows this principle is likely to be stronger than one designed by**

someone who does not.

암호시스템의 안정성은 알고리즘은 공개되고, key를 숨기는데 있다. 알고리즘이 공개된다면 많은 사람들에 의해 검증 받게 되고, 검증을 통해 암호시스템의 안정성이 확인되기 때문이다.

- **What is polyinstantiation? What is the problem of polyinstantiation in the multilevel security?**

  Polyinstantiation 이란, 데이터베이스에서 같은 Primary key를 갖는 multiple record를 허용하고, 이것을 security level로 구분하도록 하는 것이다. 즉, 데이터베이스에서 동일한 키에 대하여 다수의 튜플을 포함할 수 있도록 함으로써 낮은 수준의 사용자로부터 높은 수준의 데이터를 숨기는 데 사용되는 방지 기법이다. → 사용자의 추론을 방지!!

  Multilevel security 에서의 문제점

  서로 다른 security level 을 갖는 사용자들이 같은 primary key로 record를 만들고자 하는 경우에 낮은 등급의 사용자가 record를 기록하려는데, 해당 key를 갖는 record가 더 높은 security level에 존재 할 경우 서비스 거부의 문제가 발생하게 된다.

  반대로 높은 security level에 있는 사용자가 어떤 key로 record를 기록하려는데, 해당 key를 갖는 record가 낮은 level에 존재할 경우에는 data의 overwriting 이 발생되어 데이터 무결성이 깨지게 되고, 높은 등급의 정보가 노출될 가능성이 있다.

- **If you fly, you hear the captain say something like, "All electronic devices must switched off now, and not switched on again until I turn off the seat belt sign." Why such an announcement is necessary from the security point of view?**

  비행기 이, 착륙시 휴대용 전자 통신기기들이 전파 방해를 발생시켜 항공기의 ECU(electric Control Unit)의 오작동을 유발할 수 있기 때문이다.

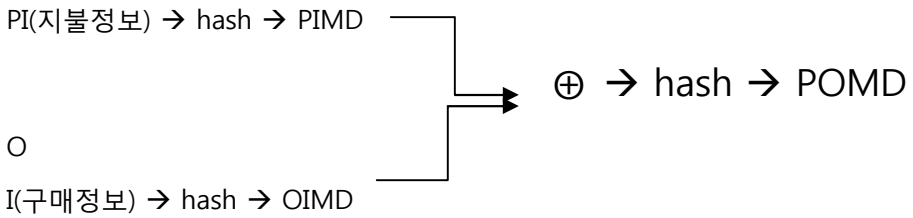
  또한 이착륙과정에서 관제탑과의 수많은 교신이 이루어지는데, 자칫 휴대전화 전파혼선으로 인해 중요한 지시를 놓치는 경우 항공기 사고로 직결될 수도 있다. 비록 휴대폰의 사용주파수와 비행기장치의 무선주파수가 달라 휴대전화 사용하는 것이 문제가 없을 것이라는 개연성은 높지만, 그래도 있을 수 있는 위험을 줄이는 노력의 일환으로 휴대전화 수신발신을 금지하는 것을 따라야한다.

- **Why dual signature is important in terms of linkability?**

두 개의 다른 메시지를 하나로 연결해서 전송하지만, 각 메시지를 개별적으로 읽을 수 있도록 하는 기술이다.

➔ 전자상거래 개인이 물건을 구매할 때 주문정보와 지불정보를 나누어 이들의 Hash값을 얻고, 이 두 값을 결합하여 또 다시 Hash 한 값을 전송하여 상점에는 주문정보만, 은행에는 지불정보만 읽을 수 있도록 한다. 이렇게 하여 불필요한 개인 정보의 노출을 막고 상점과

은행에도 데이터의 정확성과 무결성을 보장할 수 있다.



인터넷에서 거래를 할 때 구매자는 판매자를 통하여 결제 기관에 결제를 하게 된다. 이 때 판매자가 고객 정보를 통해 다른 행동을 하는 것을 막기 위하여 이중 서명을 하게 된다. 판매자는 구매자의 결제정보를 볼 수 없고 판매자가 결제 기관에 보낸 내용이 실제 구매자의 정보인지 확인이 가능하다. 이와 같이 구매자, 판매자, 결제 기관의 연계성이 있는 경우 판매자의 위,변조를 막기 위하여 이중 서명을 하게 되고 결제 기관의 경우 판매자와 구매자의 정보를 비교하여 최종 판단을 할 수 있다.

- **Explain what dictionary attack are and how to mitigate them. Explain why one-way hash and salt are used in password-based authentication.**

• Dictionary attack 이란 사전에 있는 단어, 즉 의미 있는 단어들과 그것의 대, 소문자 변형, 숫자를 추가한 값 등을 차례대로 password값으로 넣어보아 유효한 값이 나올 때까지 조사하는 방법을 말한다. 이 것을 완화시키기 위해서는 password 지정 시 의미있는 단어의 사용을 지양하고, 숫자와 문자, 대소문자의 조합을 사용하여 충분히 긴 길이의 password를 사용해야 한다.

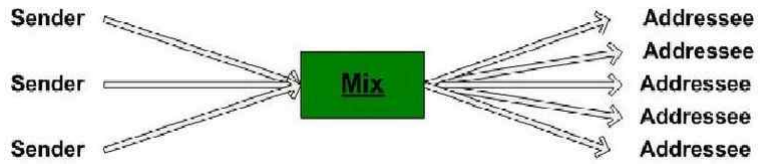
• password-based authentication 에서 one-way function 을 사용하는 것은 password를 암호화하여 저장해 놓은 정보를 갖고 역으로 password를 알아낼 수 없게 하기 위해서이다.

• salt 는 password 를 암호화 할 때 password 에 붙여서 함께 암호화하는 random 한 숫자를 말하며, 이를 통해 같은 password 를 암호화 해도 다른 결과가 나오게 되므로 안전한 password-based authentication 을 할 수 있다.

- **Explain about Chaum's mix.**

Chaum's mix는 익명성 e-mail시스템이다. 이 Mix는 발신자로부터 메일을 받아 전송하는데, 발송주소는 무작위로 anonymous하게 다량의 메일을 발송하여 발송자를 추적할 수 없도록 한다. 즉,

발송자를 추적하는데 걸리는 시간을 현실적이지 않게 만드는 것이다.



[수업 자료 단어 *형광색 시험에 출제된 문제들.. 사실상 안봐도 되긴 하지만... 보안공학완벽준비 하려면 보세요]

- **Buffer overflow attack**
- **Asymmetric information**
- **Availability**
- **Availability heuristic**
- **Backup**
- **Callback**
- **CAPTCHA**

  자동 가입 방지 코드. 어떠한 사용자가 실제 인간인지 컴퓨터 프로그램인지를 구별하기 위해 사용되는 방법으로, 인간은 구별할 수 있지만 컴퓨터는 구별하기 힘들게 의도적으로 비틀어 놓거나 그림을 주고 그 그림에 쓰여 있는 내용을 물어 보는 방법이 흔히 사용된다

- **Challenge-response protocol**
- **Cinderella attack**
- **Cognitive dissonance**
- **Confused deputy problem**
- **Cross-site script**
- **Deadlock**
- **Dictionary attack**
- **Dining philosophers' problem**
- **Evolutionary game**
- **Fallback**
- **Framing effect**
- **Integrity**
- **Lock-in**
- **Market failure**
- **Mental accounting**
- **Mixed game strategy**
- **Nash equilibrium**
- **Network externality**
- **Nonce**

- **Non-repudiation**
- **Pareto efficient**
- **Phishing**
- **Principles of least privileges**
- **Prisoner's dilemma**
- **Public goods**
- **Reference monitor**
- **Renewability**
- **Sandboxing**
- **Soft link**
- **SQL injection**
- **TOCTTOU**
- **Zero-day attack**
- **Canary**

  BOF공격을 방지하기 위해 저장되는 정보의 변형 여부를 알기 위해 canary를 스택 레이아웃에 추가한다. 포인터나 반환 주소의 변조는 이 canary value를 이용한 무결성 검사를 통해서 알 수 있게 된다.

  Stack canary is used to detect a stack buffer overflow before execution of malicious code can occur.

- **Capability**

  A capability is a ticket giving permission to a subject to have a certain type of access to an object

- **Confinement problem**
- **Confused deputy problem**

  A Confused deputy problem is a computer program that is innocently fooled by some other party into misusing its authority. It is a specific type of privilege escalation.

- **Cross-site script**
- **Principles of least privileges**

  The Principles of least privileges requires that in a particular abstraction layer of a computing environment, every module (such as a process, a user or a program on the basis of the layer we are considering) must be able to access only such information and resources that are necessary to its legitimate purpose.

- **Race condition**

  두 프로세스 간 resource사용을 위해 경쟁하는 현상

- **Reference monitor**

  A Reference monitor is a tamperproof, always-invoked, and small-enough-to-be-fully-tested-

and-analyzed module that controls all software access to data objects or devices.

- **Sandbox**
- **Script kiddie**
- **SQL injection**
- **Separation of duties**

  Separation of duties is the concept of having more than one person required to complete a task.

- **Salt**

  A salt comprises random bits that are used as one of the inputs to a key derivation function.

- **ACID**
- **Backup**
- **Battle of the forms**
- **Byzantine failure**

  Byzantine failure is an arbitrary fault that can occur during the execution of a distributed system's algorithm.

- **Callback**
- **Cinderella attack**
- **Concurrency**
- **Covert Channel**

  Covert Channel typically manipulate certain properties of the communications medium in an unexpected, unconventional, or unforeseen way in order to transmit information through the medium without detection.

- **Critical section**
- **Chinese wall model**

  Chinese wall model is a way of avoiding conflict of interest problems.

- **Deadlock**
- **Dining philosophers' problem**
- **Fallback**
- **Locking**
- **Renewability**
- **Semaphore**
- **Access Restriction**

- **Query Set Restriction**
- **Microaggregation**
- **Data Perturbation**
- **Output Perturbation**
- **Auditing**
- **Random Sampling**
- **Kerberos**

  kerberos is a computer network authentication protocol, which allows nodes communicating over a non-secure network to prove their identity to one another in a secure manner.