

전자상거래업체의 프라이버시 정책 최적화 방안 Privacy Policy Optimization for Electronic commerce

홍지명*, 박동석*, 김성집*, 장석권**
한양대학교 산업공학과*
한양대학교 경영학과**

Abstract

인터넷상의 정보 흐름에 있어서 보안과 개인의 프라이버시에 대한 중요성과 관심이 증폭되고 있다. 전자상거래가 정착되는 과정에서 정보의 통합성과 시스템에 대한 보안성을 확보할 안전한 수단은 필수적이다. 또한, 프라이버시를 보장하는 측면에서 고객정보의 저장, 이용, 노출에 대한 기업전략은 고객확보에 중요한 변수로 등장하고 있다. 본 논문에서는 프라이버시와 관련한 이슈를 고려하여 고객과 전자 상점의 관계에 대해 분석을 수행하고 전략적 차원에서 인터넷 기업의 프라이버시 정책에 대한 연구를 시도하였다. 프라이버시가 전자 상거래에 미치는 영향을 고려하여 전자 상점이 고려해야 할 적정 정보활용 모형을 개발하고 경쟁우위의 한 요소로서 신뢰도가 상거래에서 어떻게 영향을 미치는지를 분석한다.

1. 서론

인터넷은 전자 상거래의 기반 구조를 제공해 주는 중요한 가상공간으로서의 역할을 한다. 인터넷을 통한 전자 거래가 급속히 증가하여 다른 매체를 능가할 것이라는 예측이 있다. 물리적인 인프라의 급속한 발전과 함께 본격적인 전자 상거래를 예고하고있다.

인터넷의 정보흐름과 상업적 이용에 있어서 불건전한 접근, 노출 그리고 불법적 이용의 잠재적 위협이 존재하고 있다. 이러한 요소는 전자상거래 확산에 커다란 위협이 되고 있다. 고객들이 안전하게 거래할 수 있는 시스템 보안과 프라이버시에 대한 대책이 마련되어야 한다. 최근 인터넷을 통해 물품을 구입하고자 하는 사용자가 늘어나고 있지만 보안과 프라이버시에 대한 걱정은 커지고 있다. 실제로 온라인 고객의 최대의 관심사는 보안과 프라이버시이다[1].

보안과 프라이버시

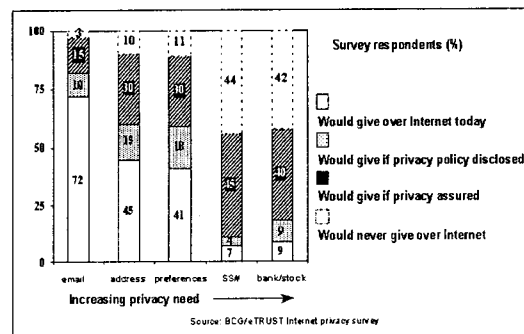
인터넷을 통한 거래에서 정보 보안에 대한 확고한 믿음이 없다면 쉽게 거래를 고려하지 않을 것이다. 전자 서류는 쉽게 복제할 수 있고 원본과 사본의 구별이 불가능하기 때문에 보안에 대한 대책은 필수적이다.[2]. 전자서명의 경우 일반 종이 문서에 행하는 서명을 디지털 문서에 공개키 암호 방식을 이용하여 수행하는 것으로 EDI/EC 분야에서 중요한 요소 기술로 등장하고 있다.

프라이버시의 개념은 본질적으로 개인 정보를 공개할 때의 이익과 그 정보를 숨기고 싶다

는 욕구 사이의 균형 점을 의미한다[3]. 전자 거래에서 기업이 고객에 대한 정보를 수집하고 활용하는 것은 관리나 분석에 있어서 필수적이다. 그러나 고객의 정보를 가진 조직이 고객의 의지와 상관없이 정보를 공개하거나 악용한다면 프라이버시 문제가 발생하게 된다. 즉, 조직은 개인의 정보를 공정하게 처리할 책임이 있고 개인은 자신에 대한 정보의 수집, 저장이나 노출에 관련한 어떤 권리를 가져야 한다.

고객과 전자상점

고객의 프라이버시에 대한 욕구는 사회가 발전하고 기호가 다양해지면서 점점 커지고 있다. 기업은 고객의 욕구를 어느 정도 충족시킬 수 있는 정책과 실행이 필요하다. 전자 상거래가 지금보다 활성화되면 프라이버시와 관련된 신뢰도가 경쟁 우위의 한 요소가 될 것이다.



<그림 1> 민감 정보성에 따른 고객의 반응

프라이버시에 대응하는 몇 가지 고객의 형태가 있다[4]. 어떤 한가지 고객정보에 대해 고객마다 느끼는 민감도의 차이라고 볼 수 있다. 이 정보에 대해 기꺼이 정보를 공개하는 경우, 프라이버시 정책을 보고 정보를 공개하는 경우, 프라이버시를 보장한다는 확신을 받고 정보를 공개하는 경우와 정보를 공개하지 않는 경우 등 4가지로 분류된다. <그림 1>은 인터넷 기업의 고객정보 요구에 대한 고객의 반응을 나타내고 있다.

현재 미국의 경우 고객과 관련된 사이트의 2/3이상이 프라이버시에 대한 언급을 하고 있다. 하지만 다양한 보안책을 제시하는 기업은 아직 적다. 기업이 공정하게 프라이버시를 공개하는지를 판단하는데 사용되는 내용은 알림(Notice), 선택(Choice), 접근(Access), 보안(Security) 등이 있다[5]. 이와 같은 내용은 프라이버시(Privacy)라는 키워드로 연결되어 있다. 고객정보는 관리, 분석 그리고 마케팅 목적으로 수집된다. 분석은 고객의 특성이나 패턴을 분석할 수 있다. 통계 분석이나 데이터 마이닝 분석을 통해 마케팅이나 의사결정에 이용한다[6].

구성

프라이버시 문제는 전자 상거래뿐만 아니라 사회 전체 분야에서 제기되는 문제다. 즉 사회 전체의 공중 도덕에 관한 문제가 된다. 본 논문에서는 인터넷상의 전자 상거래에서 기업과 고객으로 문제를 한정하고 모형화 하였다. 연구모형에서는 고객이 수집하는 정보를 정보활용수준으로 계량화하고 전자상점이 정보활용이익을 최대화하는 전략모형을 제시한다.

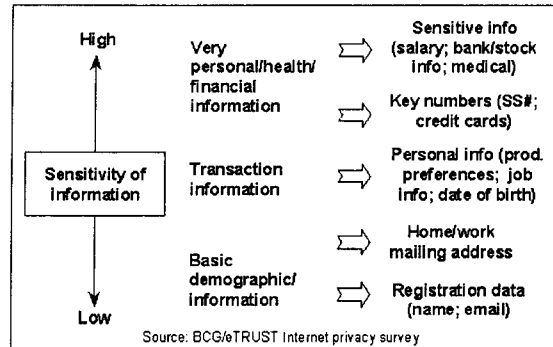
2. 연구 모형

정보의 수준에 따라 고객이 느끼는 민감도는 다르다. 이 장은 정보활용수준에 대해 정의하고 개인정보 이용도 함수를 정의하여 전자상점이 고객 수에 따른 최적 정보활용수준을 결정하는 모형을 제시한다.

고객정보

개인정보를 두 가지로 분류해보면 개인 확인정보와 통계나 기호정보로 나눌 수 있다[5]. 개인확인정보는 개인임을 확인하는 정보로 이름, 주소, 주민등록번호 등이 있다. 통계정보는 개인확인정보 외의 정보로 나이, 성, 수입, 가족사항 같은 정보이다. 또 다른 시각에서 <그림 2>처럼 민감도에 따라 정보를 분류해볼 수 있다[4]. 거래정보는 좋아하는 상품이나 취향, 거래량, 읽거나 다운로드

한 문서 등의 기호나 취향을 나타내고 사적정보는 민감하게 받아들이는 건강 신상 정보나 은행 예금 계좌나 액수 등의 정보로 나눌 수 있다. 본 논문에서는 두 번째 분류를 모형화에 이용했다.



<그림 2> 민감도에 따른 정보분류

모형화

기업이 고객정보를 수집하는 형태를 보면 한가지 특징이 있다. <표 1>에서 민감도가 적은 기본통계정보는 거의 모든 기업이 정보를 고객에게 요구하고 있다. 민감도가 높은 정보를 요구한 기업이 거의 없다. 이 데이터로 정보를 요구하는 기업의 빈도순서로 배열하면 <그림 2>에서 나타나듯이 고객이 느끼는 민감도의 순서와 거의 일치함을 알 수 있다. 신용카드정보의 경우 민감도는 높지만 그래프 상에 낮은 민감도 수준에 위치하고 있다. 물품구입 시 결재를 신용카드로 하는 사이트의 경우 즉 필수적 정보는 기본통계정보라고 가정한다.

<표 1> 고객으로부터 수집된 개인 정보

(Source: Georgetown Internet Privacy Policy Survey)

번호	항목	수량	점유율(%)	누적수량	누적점유율(%)
1	Email Address	328	19.85	328	19.85
2	Name	293	17.73	621	37.59
3	Phone Number	189	11.44	810	49.03
4	Credit Card Number	141	8.53	951	57.56
5	Age/Date of Birth	111	6.71	1062	64.28
6	Zip Code/City/State	108	6.53	1170	70.82
7	Gender	91	5.5	1261	76.33
8	Preferences/Interests	76	4.6	1337	80.93
⋮	⋮	⋮	⋮	⋮	⋮
16	Social Security Number	17	1.02	1652	100

또한 <표 2>에서 거래정보나 민감성이 높은 정보를 보유한 기업은 기본통계정보를 거의 동시에 가지고 있다[5]. 즉 기업이 보유한 고객정보의 이용도는 민감도에 따른 정보수준의 누적치라고

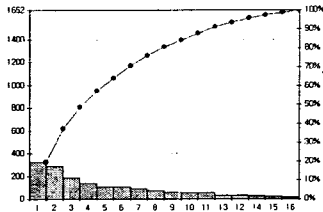
볼 수 있다.

<표 2> 고객으로부터 수집된 정보 요약

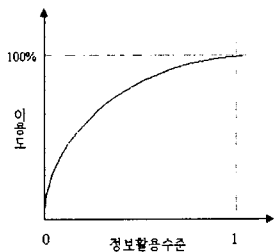
(Source: Georgetown Internet Privacy Policy Survey)

Type	Number of Sites	Percent
Personal Identifying	335	92.8%(+/-0.27)
Demographic	205	56.8%(+/-0.051)
Personal Identifying Only	132	36.6%(+/-0.050)
Demographic only	2	0.6%(+/-0.008)
Both Personal Identifying & Demographic	203	56.2%(+/-0.051)
None	24	6.6%(+/-0.025)
Total	361	100%

즉 <그림 3>에서 X축은 고객정보 활용 수준을 나타내고 Y축은 선과 같은 누적된 정보수준을 보유한 기업 빈도수의 합이다.



<그림 3> 표 1에 대한 파레토분석



<그림 4> 고객 1인당 개인정보 이용도

고객이 느끼는 민감도의 항목을 0과 1사이의 선상에 배열한다면 그림 4에서의 X축처럼 표시될 수 있다. Y축은 기업이 보유한 정보수준의 이용도라고 정의한다. 기업이 높은 정보수준을 보유하면 그 만큼 이용도는 커지게 된다. 기업이 보유한 고객정보의 이용도가 그림 4와 같이 된다면 그 그래프는 다음과 같은 식 (1)로 모형화 된다.

$$g(x) = x^c \quad (0 < c < 1, 0 < x < 1) \quad (1)$$

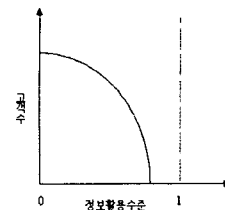
c: 서비스 차별계수

식 (1)에서 c는 서비스 차별계수로 정의

한다. 같은 정보항목이라도 전자 상거래내의 각 산업마다 이용도 함수는 다를 수 있다.

전자상거래내의 한 산업이 어떤 수준의 정보활용수준을 가지면 고객마다 수집된 정보의 이용도는 정해진다. 일인당 고객의 이용도가 정해지면 기업이 고객 한 사람으로부터 얻을 수 있는 정보활용이익이 결정된다고 볼 수 있고 많은 고객을 확보한 기업은 그만큼 정보활용이익은 크다. 하지만 정보활용수준을 올리면 정보활용이익이 증가하지만 정보활용수준을 올릴수록 고객이 느끼는 민감도는 커져 고객의 수는 감소할 것이다. 기업의 정보활용이익에 있어서 고객 수와 정보활용수준은 교환(Tradeoff)관계임을 알 수 있다.

기업이 요구하는 정보활용수준에 고객의 반응이 어떠한지를 알기 위해 설문조사를 실시했다. 기업이 정한 특정 정보활용수준에 따른 고객수에 대한 추세를 추정할 수 있기 때문이다. 조사 결과를 고객 수에 대한 빈도수로 표시하면 <그림 5>와 같은 패턴임을 알 수 있다. 고객이 느끼는 민감도가 높을수록 정보제공을 급격히 꺼리는 형태로 반응하고 있다.



<그림 5> 정보활용수준에 따른 고객 수

Y축은 <그림 4>와 동일한 X축에서 기업의 정보활용수준에 따라 고객수의 반응을 나타내고 있다. 이를 모형화 하면 다음과 같은 식을 얻을 수 있다.

$$f(x) = b - ax^2 \quad (a > 0, b > 0) \quad (2)$$

a: 신뢰 계수

b: 초기 상태의 고객 수

식 (2)에서 b는 기업이 개인정보 요구하지 않았을 때의 고객 수를 나타내고 a는 신뢰 계수로 기업의 신뢰 수준을 의미한다. 기업이 어떤 정보활용수준을 정하면 고객 일인당 이용도와 고객수가 정해진다. 이때 이용도와 고객 수를 곱하면 기업이 활용하는 전체고객의 총 이용도가 다음과 같이 산출된다.

$$V(x) = f(x) * g(x) = b x^c - a x^{2+c} \quad (3)$$

3. 분석

기업은 고객의 이용도를 활용하여 고객에 대한 정보활용이익을 얻는다. 전체고객의 총 이용도가 크면 그만큼 고객으로부터 얻는 정보활용이익은 증가한다고 가정한다. 위 총 이용도 식을 두 번 미분하면 다음과 같이 된다.

$$\frac{d^2 V(x)}{dx^2} = c(c-1)bx^{c-2} - a(2+c)(c+1)x^c \quad (4)$$

식 (4)에서 (c-1)이 0보다 작기 때문에 총 이용도 함수는 0보다 적다. 결국 식 (3)은 오목 함수(Concave)이고 최적 정보활용수준이 존재한다. 총 이용도 함수의 최적 정보활용수준은 다음과 같이 유도된다.

$$\frac{dV(x)}{dx} = x^{c-1}[cb - a(2+c)x^2] = 0$$

로부터

$$x^* = \sqrt{\frac{cb}{a(2+c)}} \quad (5)$$

식 (5)에서 기업은 총 이용도를 최대화시키는 최적 정보활용수준을 택하는 전략을 사용할 것이다. 총 이용도를 최대화시키면 정보활용이익이 커지고 기업의 비용적 이익도 증가할 것이다.

민감도 분석을 수행하기 위해 식 (2)의 신뢰 계수를 변형하여 신뢰도 지수 α 를 정의하면 다음과 같다.

$$\frac{1}{a} = \alpha \quad (6)$$

이 신뢰도 지수의 값이 클수록 기업의 프라이버시정책에 대한 고객의 믿음은 커진다. 식 (5)의 정보활용수준을 목적함수인 식 (3)에 대입하면 다음과 같다.

$$V(x^*) = 2 a c^{\frac{c}{2}} \left(\frac{b}{2+c}\right)^{\frac{c}{2}+1} \quad (7)$$

위 모형을 통해 프라이버시에 대한 신뢰도가 기업의 경쟁 우위에 어떠한 영향을 주는지 추론해볼 수 있다.

- 고객의 정보활용을 높일수록 거래하는 고객 수는 급속히 감소한다. 식 (2)에서 x 축 상의 값이 증가할수록 즉 $\angle x$ 에 따른 고객 수는 급격히 감소한다. 민감도가 높은 정보일수록 고객은 정보제공을 꺼리고 있다.
- 신뢰도가 낮은 기업일수록 정보활용수준은 낮아진다. 식 (2)를 보면 고객확보 측면에서 α 지수가 작은 기업일수록 경쟁력 있는 기업보다 상대적으로 x 축 상의 값은 적어진다.
- 신뢰도가 높은 기업은 그렇지 않은 기업보다 전체 고객의 총 이용도는 커진다. 식 (7)에서 b 값을 고정하면 α 가 큰 기업일수록 그렇지 않은 기업보다 총 이용도 값은 크다.

4. 결론 및 과제

이후 과제로 전자 상거래내의 다양한 서비스에 따른 실증연구를 수행하여 a, b, c 에 대한 값을 추정하여 전자상점이 선택 가능한 적정 프라이버시 수준을 제시하고 프라이버시 전략에 따른 전자상점의 비용분석도 흥미 있는 연구가 될 것이다.

본 연구의 결과를 보면 전자상거래에 있어서 신뢰성 있는 기업일수록 고객 수를 많이 확보하고 얻는 이익도 크다는 것을 알 수 있다. 고객이 신뢰성 있는 상점을 선택하는 것은 사업자간의 경쟁을 유발하고 결국 프라이버시 보호 수준은 더욱 높아질 것이다.

참고문헌

1. NetZero Report, "Privacy Survey Results" August 10 1999.
2. B. Clifford Neuman, Security, Payment, and Privacy for Network Commerce IEEE Journal of Selected Areas in Communications, Vol. 13, No. 8, October 1995.
3. Dave Redell, Information Technology and the Privacy of the Individual, Daft ACM Whitepaper on computer and privacy, September 1992.
4. eTrust Internet Privacy Study, The Boston Consulting Group, <http://www.bcg.com>.
5. Georgetown Internet Privacy Policy Study, a report to the Federal Commission; March 1999.
6. 홍지명, 김성집, 데이터 마이닝에서 샘플링 기법을 이용한 연속 패턴 알고리즘, Volum. 21/45 공업경영학회, Feb. 1998.