Trinh, Tuan Anh; Gyarmati, Laszlo

Conference Paper

Privacy driven internet ecosystem

23rd European Regional Conference of the International Telecommunications Society (ITS), Vienna, Austria, 1st-4th July, 2012

Provided in Cooperation with:
International Telecommunications Society (ITS)

Suggested Citation: Trinh, Tuan Anh; Gyarmati, Laszlo (2012) : Privacy driven internet ecosystem, 23rd European Regional Conference of the International Telecommunications Society (ITS), Vienna, Austria, 1st-4th July, 2012, International Telecommunications Society (ITS), Calgary

This Version is available at:
https://hdl.handle.net/10419/60351

Privacy Driven Internet Ecosystem

Tuan Anh Trinh
Department of Telecommunication and Media Informatics
Budapest University of Technology and Economics
Magyar tudosok krt. 2,
H-1117, Budapest, Hungary
trinh@tmit.bme.hu

Laszlo Gyarmati
Telefonica Research
Plaza de Ernest Lluch i Martin, 5, floor 15
08019, Barcelona, Spain
gyarmati@tid.es

ABSTRACT

The dominant business model of today's Internet is built upon advertisements; users can access Internet services while the providers show ads to them. Although significant efforts have been made to model and analyze the economic aspects of this ecosystem, the heart of the current status quo, namely privacy, has not received the attention of the research community yet. Accordingly, we propose an economic model of the privacy driven Internet ecosystem where privacy is handled as an asset that can be traded. Expressing the entropy of privacy as the service providers' fitness value and applying a dynamic network formation model based on the preferential attachment principle allow the analysis of the providers' economic interactions in a realistic framework, whose properties are illustrated based on extensive simulations.

1. INTRODUCTION

Along with the development of Internet services, several technical terms have been widely accepted by the research community and the society as well. The multimedia services give rise to the term quality of service (QoS) that has been transformed to quality of experience (QoE). Numerous technical as well as economic models have been proposed in the field of pricing Internet access [7, 10], QoS [6, 3], and QoE [8, 9]; these works contributed to the understanding of the different aspects of these domains. In addition, the economic modeling and analysis formed the whole ecosystem of Internet based services.

The next wave of Internet services, identified with the flagships applications like online social networks, user generated content sharing, established a novel business model in which the services can be used free of charge—at the first sight. The Internet is not operated on the principle of charity, on the contrary, the profit awareness of the whole ecosystem is continuously augmenting. As a sign of Internet's maturity, the market determines the viability of the newly introduced services rapidly: a service is either profitable or it will soon vanish from the service map of the Internet.

The heart of the current ecosystem is advertisement; the payment of advertisers covers the expenditures of the content providers allowing them to offer the services for free. Beneath the surface, the users' privacy is exchanged for the advertisers' money. Accordingly, privacy can be treated as the currency of this economic interaction.

From this point of view, it is essential to understand, model, and analyze the economics of privacy in order to develop an applicable framework for the Internet ecosystem for the forthcoming years. The economics of privacy research area includes several main topics including deriving viable business models for privacy as a service, pricing for privacy, and pricing of privacy. We focus on the later one; accordingly, in this paper we propose a framework that can model the privacy driven ecosystem of the Internet. To our best knowledge, our work is the first of its kind, i.e., the pricing of privacy issue has not been addressed yet. The main contribution of our work is threefold:

- we claim that privacy should be handled as an asset, which underpins the current Internet ecosystem;
- we propose a privacy trading Internet ecosystem framework, incorporating privacy as a fitness value into Barabasi's model dealing with the dynamics of companies;
- we present extensive simulation results to gain insight about the driving forces of such an ecosystem.

The structure of the paper is as follow. First, we introduce our model in Section 2; the modeling of privacy, service providers, and privacy trading will be revealed. To give an insight about the presented framework, Section 3 presents an example where the privacy driven ecosystem of an imaginary market is demonstrated. Simulation results are shown in Section 4 to quantify the impact of the model's parameters. After a discussion (Section 5), the paper concludes in Section 6.

2. MODEL

In this section, we propose a novel model for the privacy driven Internet ecosystem. Specifically, we model the pricing of privacy, i.e., how can the users' privacy be handled as an asset.

2.1 Privacy

As we have introduced earlier, we believe that privacy underpins the Internet ecosystem, i.e., the advertisement industry utilizes it to display appropriate ads to the users. Therefore, we start our framework's description by modeling privacy as an asset. Privacy is a collection of personal information that describes the users' behavior. Accordingly, it is straightforward that the value of a privacy dataset can be expressed quantitatively. By applying a quantitative metric, privacy can be considered as an asset, similar to other assets like currencies, securities, real estates, etc. For simplicity reasons, we apply a one-dimensional metric for assessing privacy. We note however that due to the complexity of a privacy dataset, multi-dimensional metrics can also be created, which may describe the value of the privacy dataset more accurately.

Information theory provides a rich and well-established mathematical framework to quantify information. *Entropy* is one of the most crucial principle of information theory; the concept introduced by Shannon [11] quantifies the value of information, usually in bits. Therefore, entropy offers itself to be used as a privacy metric in or framework.

With the use of entropy, we are able to quantify the value of a privacy dataset. However, the definition of privacy does not take into account the fact that the value of information depends on the time as well. Considering the privacy assets of service providers, the impacts of time cannot be disregarded, i.e., a user's behavior worth more today as it would worth in a year or two. Thus, we compute the value of privacy by incorporating these two factors such as entropy and time. The formal definition of the value of a privacy asset is

$$P = f(E_1, E_2, \ldots, E_k, t_1, t_2, \ldots, t_k, T) \qquad (1)$$

where the service provider has k privacy datasets with E_i entropy and t_i timestamp, while the actual time is denoted by T. The actual characteristic of the function may highly depend on the nature of the service providers' business; in the simulations we will apply a linear function, which captures the time dependency of privacy's value:

$$f(E_1, \ldots, E_k, t_1, \ldots, t_k, T) = \frac{E_1}{T+1-t_1} + \cdots + \frac{T_k}{T+1-t_k} \qquad (2)$$

2.2 Service providers

In terms of the stakeholders, service providers are included to the model; thus, they can interact with each other to trade privacy. The service providers are modeling the companies of the Internet, from the local Internet Service Providers to the operators of the Web 2.0 services. Indeed, all these companies possess some amount of privacy, out of which they can produce revenues based by applying different business models. A service provider can be described with three properties: monetary asset, privacy asset, and privacy production. First, the monetary asset, denoted by M, means the amount of money that the service provider owns. The provider's monetary asset alters if the provider is involved in a privacy transaction, i.e., sells or buys privacy. Second, the privacy asset, denoted by P, of the provider incorporates all the privacy datasets that the provider has. A privacy dataset has an additional descriptor along with the abovementioned two aspects (entropy and time stamp): the origin of the privacy dataset. The service provider can have privacy generated by its own users and can buy privacy datasets from the other service providers. The third property of service providers is related to this topic; a service provider may produce privacy.

The service providers are profit-oriented companies; accordingly, they realize revenues and have to pay the costs of their services. The origin of the providers' revenue can be twofold. First, based on their privacy datasets service providers make money in each time interval, e.g., in every month. The amount of the revenue is a function of the actual valuation of provider's privacy asset. Clearly, there is a strong correlation between the revenues generated by the provider and the value of its privacy asset: more incomes can be generated based on an up-to-data privacy dataset. The roundly generated revenue of the service providers is formalized as:

$$\hat{M} = g(f(E_1, \ldots, E_k, t_1, \ldots, t_k, T)) \qquad (3)$$

The nature of the g function can be highly diverse among the service providers in practice, e.g., it may reflect the strategic decisions of the providers like the number of advertisements displayed to a single user. We apply a single linear revenue model in our simulations; thus, the revenue of the service providers can be formalized as:

$$g(f(E_1, \ldots, E_k, t_1, \ldots, t_k, T)) = \alpha f(E_1, \ldots, E_k, t_1, \ldots, t_k, T) \qquad (4)$$

where α is a revenue constant. The second source of the revenues is selling privacy on the market; this aspect will be covered in details shortly.

From the expenditures point of view, the service providers face various types of costs owed to their operation. Although a complex model incorporating these factors may describe the costs more realistically, for simplicity reasons we assume in our model that the service providers' expenditures are proportional to their the volume of their privacy assets. If the linear constant

is denoted by β, the roundly cost of a service provider is formalized as:

$$c = \beta \sum_{i=1}^{k} E_k \tag{5}$$

Analogous to the revenue constant, β may also company specific in a general case reflecting the efficiency of the companies business processes.

Based on the revenues and expenditures, the monetary asset of the services providers can be computed in every round as:

$$\hat{\Pi} = \Pi + \hat{M} - c \tag{6}$$

In case of the simulations, the monetary assets of the service providers are updated periodically.

Some service providers, whose services are used by a number of persons, can produce privacy. As the dynamics of the users is out of the scope of our model, we assume that the entropy of the privacy datasets is fixed in the Internet ecosystem. Therefore, we model the produce of privacy throughout the timestamps of the privacies. Particularly, if a service provider is capable to produce some amount of privacy, the timestamps of these privacy datasets are synchronized periodically to the actual time. This way the value of these privacies is maintained.

2.3 Trading privacy

In our proposed framework privacy is considered as an asset; accordingly, privacy can be traded among the service providers. Similar to other transactions, there exist a buyer and a seller in case of a transaction. The buyer service provider pays a certain amount of monetary asset to the seller provider. Albeit two kind of service providers, the ones that can produce privacy and those who cannot, are modeled in the proposed framework, any company can sell a portion of its privacy asset and thus realize additional revenues.

In case of a privacy trade, the following events occur. First, a privacy demand arises at one of the service providers, i.e., it wants to buy a certain amount of privacy denoted by p. This privacy claim should be supported with the sufficient amount of money; the price of the privacy has to be paid. For simplicity reasons, we assume that the price of the privacy is proportional to the its volume:

$$s = h(p) = \gamma p \tag{7}$$

where γ denotes the selling constant. The actual value of γ may depend on several factors; however, obviously a price should be at least as much as the revenue decrease of the seller.

Due to the nature of information, the value of a privacy dataset is reduced if more than one service providers have an access to it. Unlike material assets like cars, houses, or stocks of companies, information can be copied practically without costs, allowing multiple entities to own it. Accordingly, if a service provider sells p privacy the value of its privacy asset will be decreased. In particular, the privacy reduction of the seller is $t(p)$; any arbitrary function may be used to describe this lost of privacy. To handle the impact of the sold privacy along with the other privacies of the provider, we add the sold privacy with a negative entropy value to the privacy assets of the provider. The timestamp of the sold privacy is the time of the transaction. Similarly, the timestamp of the buyer's privacy asset is set to the time of the transaction because this information is novel to the buyer. Formally, the seller will have a new virtual privacy asset with the $-p, t_{\text{transaction}}, 0$ properties. As a result, the impact of the privacy selling decreases as the time passes analogous to the values of the ordinary privacy datasets. Therefore, the seller's privacy value is computed as follow:

$$f(E_i, -p, t_i, t_p, T) = \sum_{i=1}^{k} \frac{E_i}{T + 1 - t_i} - \frac{p}{T + 1 - t_p} \tag{8}$$

As the amount of handled privacy is not decreased due to the privacy selling, the expenditure of the providers is computed only based on the privacy values with positive entropy.

To summarize the details of a privacy transaction, we present the change of the monetary and privacy assets both in case of the buyer and seller service providers:

$$\hat{P}_{\text{buyer}} = P_{\text{buyer}} + p \tag{9}$$

$$\hat{M}_{\text{buyer}} = M_{\text{buyer}} - h(p) \tag{10}$$

$$\hat{P}_{\text{seller}} = P_{\text{seller}} - t(p) \tag{11}$$

$$\hat{M}_{\text{seller}} = M_{\text{seller}} + h(p) \tag{12}$$

The last missing piece of the privacy driven Internet ecosystem is how are the parties of a privacy trading transaction selected. On the one hand, the demands for privacy are arising based on a given probability distribution, i.e., one of the service providers wants to buy some privacy asset time after time. Contrary to this, it is not straightforward which service provider should sell the required privacy. Recall that we assume a perfect market in terms of the uniform prices on the market. Therefore, the identity of the seller should be determined based on other mechanisms.

Several requirements exist towards the applied seller determining mechanism such as

- the service providers have different amount of privacy assets which should be reflected by the method;
- human decisions are usually not deterministic even if there are rational aspects;
- every service provider deserves the opportunity to

be selected as a seller if it owns the required amount of privacy; however, successful stakeholders are usually more preferred.

The introduced requirements are in accordance with the properties of the network model of Barabasi [1], where the network of the entities is grown based on the preferential attachment model (also known as the richer gets richer principle). To be applicable in case of our privacy driven Internet ecosystem framework, the original Barabasi model has to be adapted. One of the several extensions of the scale-free network generation model is the introduction of fitness values [2]. In this case, the new connection between the entities is formed based on the relative fitness values, i.e., a company with larger fitness receives more links than the others.

The privacy ecosystem is quite similar to the market of the search engines; thus, we propose an adaption of the Barabasi model to be used in our framework. In case of the Internet ecosystem, the network's nodes represent the service providers. Links are formed between two providers if they are involved in a privacy trading transaction, namely, one service provider buys p amount of privacy from an other service provider. As the goal of the transactions is to acquire privacy information, we consider the value of the privacy owned by a provider as the fitness of that service provider. Thus, the preferential attachment principle is applied using the providers' owned privacy values.

Accordingly, the identity of the privacy seller is determined as follow. First, the buyer expresses its demand indicating the required volume of privacy (p). Next, those service providers are selected whose privacy's value is at least as large as the required volume ($P \geq p$). Afterwards, the seller is selected among these providers based on the preferential attachment principle. Let us assume that there are m possible providers that can sell p amount of privacy; the service providers are selected with the following probabilities:

$$\text{Prob}_1 : \ldots : \text{Prob}_m = P_1 : \ldots : P_m \qquad (13)$$

The proposed privacy trading framework can be used to model and to analyze the privacy driven Internet ecosystem; in the forthcoming sections we illustrate the applicability of the framework.

3. ILLUSTRATIVE EXAMPLE

Before presenting the simulation results, we show a simple example to illustrate the concepts of our privacy pricing framework. There are four service providers on the imaginary market; each of them has diverse parameters. The first one is the largest, it has $M_1 = 1000$ unit of money, it has a single privacy dataset with $E_1 = 300$ as entropy and $t_1 = 1$ as timestamp, and it produces the privacy itself. The second provider owns $M_2 = 500$ money, a privacy with $E_2 = 150$ entropy and $t_2 = 1$ timestamp, and it is capable to produce privacy. The third service provider has $M_3 = 200$, $E_3 = 100$ amount of privacy with $t_3 = 2$ timestamp, and it can also create privacy. The last provider cannot produce privacy, it does not have privacy asset at the beginning; however, it has the largest monetary asset ($M_4 = 5000$). The properties of the providers as well as the roundly amount of their money and value of their privacy is shown in Table 1.

	1	2	3	4
money	1000	500	200	5000
privacy	300,1,1	150,1,1	100,2,1	0,0,0
T=2	1000,150	500,75	200,100	5000,0
T=3	950,100	475,50	200,50	5000,0
T=4	875,75	438,38	183,33	5000,0
T=5	privacy update			
	1025,300	513,150	233,100	5000,0
	4 buys 100 from 1			
T=6	1425,50	513,75	233,50	4550,100
T=7	1325,50	488,50	217,33	4550,50
T=8	privacy update			
	1542,267	588,150	283,100	4533,33

Table 1: Illustrating example of the privacy trading framework. The presented rounded values are money and privacy; respectively.

Next, we walk through a few round and compute the assets of the providers to highlight the properties of the proposed framework. Throughout the example we assume that the revenue constant is $\alpha = 1$, the cost function is $\beta = 0.5$, the selling constant is $\gamma = 5$, while the privacies are updated in every third rounds.

Our example starts at $T = 3$, based on the introduced parameters of the service providers we can compute the value of their monetary and privacy asset before the start. The value of the first provider's privacy is $300/3 = 100$; based on this its money can be computed as $1000 + 100 - 0.5 * 300 = 950$ where the terms denote the money of the previous round, the actual revenue based on the update privacy value, and the costs. The second provider has $150/3 = 50$ privacy value and $500 + 50 - 0.5 * 150 = 475$ as money. The value of the third provider's privacy is computed slightly differently resulting from the diverse timestamp; it is $100/2 = 50$ while its money is $200 + 50 - 0.5 * 100 = 200$. As the last provider does not have privacy asset, its monetary asset does not alter. The actual value of the assets can be computed analogously for the next round ($T = 4$).

Before the fifth round the privacies are updated; therefore, their value is much larger than they were in the former round. Accordingly, the privacy values are identical with the entropies as a result of the selected revenue constant: 300,150,100; respectively. At the end of this round the fourth provider wants to buy 100 amount

of privacy; the price of this is 500 unit of money. As all the other providers have at least 100 as their privacies; the seller is picked out of the three providers based on the preferential attachment. Because the most possible seller (54.5%) is the first provider, let us assume that it is picked. Afterwards, the buying price is transferred to the seller. In the next round ($T = 6$), the impacts of the transaction are realized, i.e., the privacy value of the first provider is computed considering the privacy loss caused by the sell: $150 - 100 = 50$. On the other hand, the fourth provider will have 100 as its privacy value.

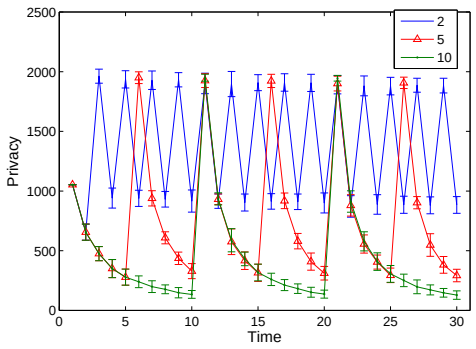
In the next rounds, the process of computing the privacy and monetary assets is similar to the presented ones. The last round of the example ($T = 8$) shows the impact of privacy produce; the all the privacy values are updated except that of the fourth provider. Thus, it will have only 33.3 as privacy asset.

4. SIMULATION RESULTS

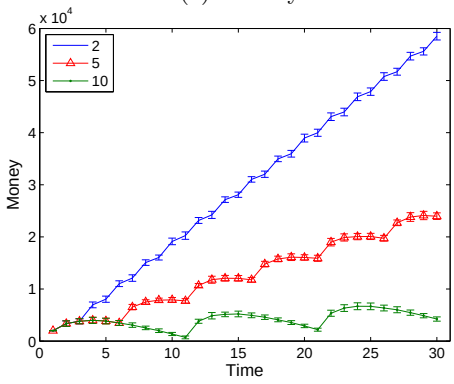
This section highlights some of the characteristic of the proposed framework based on extensive simulations. The presented rules are incorporated into a discrete-event simulator, which is capable to model the privacy interactions on a service provider market. The source code of the simulator is made publicly available at [5]. The presented results are averaged over 50 simulation runs, the default parameters are $\alpha = 2$, $\beta = 0.5$, $\gamma = 5$, the privacy buying demands (**Exp**(1)) and the demands' size (**Exp**(0.01)) are generated based on exponential distributions, while the privacies are updated in every fifth round.

Figure 1 quantifies the impact of the privacy update periods on the privacy and money of a service provider capable to generate privacy. At the beginning of the simulation, the provider has $E = 1000$ privacy asset. If the privacy is updated frequently, i.e. in every second round, the value of the privacy asset is oscillating between two values (Figure 1(a)). The value of the privacy is decreased first; however, in the next round the privacy is updated resulting high value again. The larger the privacy update period the lower the value of the privacy asset becomes. The fast depreciation of the privacy's value is caused by the applied value function (Equation 2). In terms of the money, the service provider has a larger amount of money in case of more frequent privacy updates (Figure 1(b)). Again, the waves of the plot are a result of the used value function. The implication of the results is that providers should develop services that involve users' participation frequently; thus, the privacy asset can be updated often, which at the end results increase revenues.

Next, we analyze how the inter-demand parameter impacts the assets of a service provider (Figure 2), which cannot produce privacy; thus, the provider buys privacy from the other stakeholders. Figure 2(a) shows the value of the provider's privacy in case of several privacy trading frequencies. If the inter-demand parameter is low, a small amount of privacy is bought in every round; accordingly, the privacy asset is increasing slowly. As the parameter augments, the provider buys privacy chunks more often [1], causing larger and larger privacy assets. Buying privacy is costly; therefore, the money of the provider runs out soon (Figure 2(b)). Without money, the provider cannot buy more privacy then the value of its privacy asset starts to decrease. The results reveal that a provider cannot be successful only by buying large privacy dataset; it should utilize the privacy information to generate revenue, which can assure its long-term profitability.



(a) Privacy



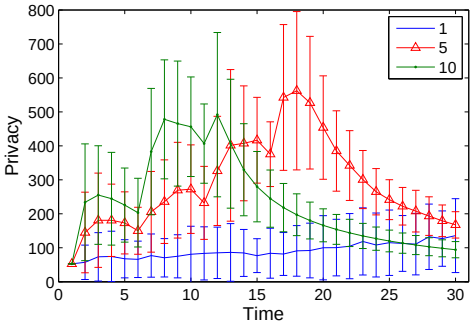
(b) Money

Figure 1: The impact of the privacy update period in case of a producer service provider

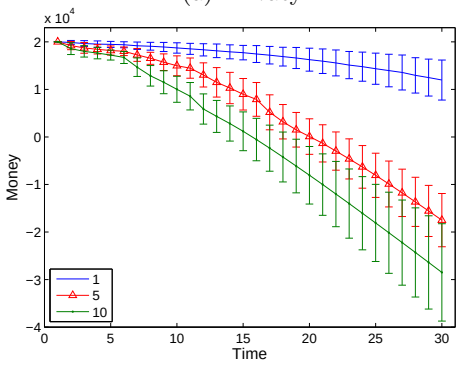
5. DISCUSSION

The presented privacy driven framework incorporates all the stakeholders of the Internet ecosystem. Throughout the paper we assumed that the popularity of the services is constant; thus, the users do not start and terminate using a specific service. However, the framework can be extended to include to end-users as well. To this

[1]Recall that the inter-demand time is generated randomly using exponential distribution; thus, the expected value of intervals is the reciprocal of the parameter.



(a) Privacy



(b) Money

Figure 2: The impact of the inter-demand period in case of a privacy buyer service provider

extent, an additional modeling layer should be created, where the dynamics of the users is considered. The users tend to use the most innovative services; thus, the Barabasi model can be used again with the providers' innovativeness as a fitness value. For example, if a novel, highly innovative service appears on the market the users will migrate to the new provider. As more users are using the service more and more privacy is generated, which can be sold on the service providers' market. Therefore, the same privacy—in terms of the group of users—is traded in the network of the providers; however, the direction of the privacy flow will be altered.

The proposed framework describes the value of privacy with its entropy, which should be determined for a given dataset to trade privacy. Computing the entropy of a large dataset may seem challenging at a first sight, but there exist methods that address this issue like [12]; moreover, a recently proposed method [4] is able to determine the entropy of large dataset as fast that the method is applicable even in real-time systems.

The current model assumes that the providers are always willing to sell their privacy to the others. Considering the competitive nature of the Internet ecosystem, this may not be valid in a general case because providers would behave strategically estimating the proper selling price. The game theoretic analysis of such a strategic framework seems to be a challenging research topic.

6. CONCLUSION

In this paper, the network economics of privacy is covered. In particular, we proposed a framework to model the today's privacy driven Internet ecosystem. At the heart of the model privacy is handled as an asset, which can be traded among the market's stakeholders. Entropy based privacy evaluation and an extension of preferential attachment, using privacy as the fitness of the service providers, result a dynamic description of the privacy trading. We believe pricing of privacy will be a justifiable successor of the efforts dealing with pricing of QoS and QoE. Therefore, we hope that the presented results will encourage further discussion and research in the area of the network economics of privacy.

7. REFERENCES

[1] A.-L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, 1999.

[2] G. Bianconi and A. Barabási. Competition and multiscaling in evolving networks. *EPL (Europhysics Letters)*, 54:436, 2001.

[3] L. DaSilva. Pricing for QoS-enabled networks: A survey. *Communications Surveys & Tutorials, IEEE*, 3(2):2–8, 2000.

[4] P. Li. Compressed Counting and Application in Estimating Entropy of Data Steams. In *Workshop on Algorithms for Modern Massive Data Sets (MMDS)*, 2010.

[5] Network Economics Group. Privacy Driven Internet Ecosystem Simulator. `http://netecon_group.tmit.bme.hu/source-codes`.

[6] A. Odlyzko. Paris Metro Pricing: The minimalist differentiated services solution. In *IWQoS'99*, pages 159–161. IEEE, 1999.

[7] A. Odlyzko. Internet pricing and the history of communications. *Computer Networks*, 36(5-6):493–517, 2001.

[8] P. Reichl. From Charging for QoS to Charging for QoE: Internet Economics in the Era of Next Generation Multimedia Networks. pages 231–231, 2007.

[9] P. Reichl, B. Tuffin, and R. Schatz. Economics of logarithmic Quality-of-Experience in communication networks. In *CTTE 2010*, pages 1–8. IEEE, 2010.

[10] S. Shakkottai and R. Srikant. Economics of network pricing with multiple ISPs. *IEEE/ACM Transactions on Networking (TON)*, 14(6):1233–1245, 2006.

[11] C. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27:379–423, 1948.

[12] H. C. Zhao, A. Lall, M. Ogihara, O. Spatscheck, J. Wang, and J. Xu. A Data Streaming Algorithm for Estimating Entropies of OD Flows. In *Proc. IMC'07*, pages 279–290, New York, NY, USA, 2007. ACM.