Mobile Networks and Applications 7, 213–223, 2002

Defending Wireless Infrastructure Against the Challenge of DDoS Attacks *

XIANJUN GENG, YUN HUANG and ANDREW B. WHINSTON **
Center for Research in Electronic Commerce, Department of MSIS, McCombs School of Business, The University of Texas at Austin, Austin, TX 78712, USA

Abstract. This paper addresses possible Distributed Denial-of-Service (DDoS) attacks toward the wireless Internet including the Wireless Extended Internet, the Wireless Portal Network, and the Wireless Ad Hoc network. We propose a conceptual model for defending against DDoS attacks on the wireless Internet, which incorporates both cooperative technological solutions and economic incentive mechanisms built on usage-based fees. Cost-effectiveness is also addressed through an illustrative implementation scheme using Policy Based Networking (PBN). By investigating both technological and economic difficulties in defense of DDoS attacks which have plagued the wired Internet, our aim here is to foster further development of wireless Internet infrastructure as a more secure and efficient platform for mobile commerce.

Keywords: DDoS attack, wireless infrastructure, PBN, Wireless Extended Internet, Wireless Portal Network, Wireless Ad Hoc Network

1. Introduction

The wireless Internet has become an exciting realm for m-commerce at an amazing speed. The estimated number of wireless subscribers was 109 million in December 2000 in the US alone, according to a semi-annual wireless industry survey conducted by Cellular Telecommunications Industry Association [4]. It represented an increase of 27.2% from a year earlier, adding nearly 23.43 million new users. According to a new study released by Strategy Analytics, the global cellular market will grow at an annual rate of 17% over the next five years, reaching $700 billion with 1.4 billion global wireless subscribers by 2005 [22].

M-commerce is not a simple duplication of e-commerce upon wireless devices. As pointed out by market research institutions including Goldman Sachs [10] and Bear Stearns [1], "m-commerce is about information and transactions that are timely" [1, pp. 140].

Is wireless infrastructure ready for time-sensitive m-commerce? From a technological perspective, it is ready for anytime, anywhere access. 3G wireless technology also enables high-speed access. However from a security perspective, time-sensitive m-commerce is vulnerable to network delays or even network denial caused by a dangerous type of security problem – the Distributed Denial-of-Service (DDoS) attack – that has been much publicized but seldom understood completely [9,16].

Due to the time-sensitive nature of m-commerce, it is not surprising for wireless infrastructure providers to carefully plan the radio spectrum allocation and pricing to avoid any predictable congestion. Given the huge cost of radio spectrum rights, they also have enough incentive to defend against most security risks through constant and prompt patching of system security holes and real-time monitoring. These remedies, however, target unauthorized intrusions. A DDoS attack, on the other hand, never tries to break into the victim's system. On the wired Internet, attacks against well-known sites [8,12,13] have repeatedly proved the lack of an effective defense. As Geng and Whinston [9] pointed out, effective defense is unlikely to appear on the present wired Internet as there lacks an incentive structure to push cooperation on the wired Internet.

DDoS attacks are not a serious problem to the current wireless Internet, in part because of the extremely limited and often non-programmable functionalities of current mobile devices. However, our research strongly suggests that DDoS attacks can be a real threat in the near future given the increasing computational power, network bandwidth, and users in the wireless Internet economy. Two significant events have already occurred. First, in the summer of 2000, there appeared the first preliminary virus against mobile phones [6]. Furthermore, Eugene Kaspersky, head of anti-virus research at Kaspersky Lab, a Moscow-based anti-virus company, once commented on this virus [17]:

> "This is not the first and obviously not the last security breach discovered in mobile phones. Moreover, I believe as more functionality is added to mobile phones, it will result in more breaches being found."

The second event was the emergence of the first DDoS attack tool toward mobile phones, known as the SMS-flooder [21]. It tries to use the wired Internet to attack a wireless victim. First it proliferates through Microsoft Outlook just as the Melissa virus (see `http://www.cert.org/advisories/CA-1999-04.html` for details) does. Then it commands all infected Microsoft Outlook software to send short messages (SMS-messages) to a certain victim's mobile phone to inundate it. The potential hazard is not only in the blocking of communications but also in the high financial cost if pricing is usage-based.

* Research supported in part by Intel.
** Corresponding author.

The two events mentioned above show that the DDoS attack directed towards the wireless Internet is not only a theoretical possibility, but also a real and evolving threat. However, research is lacking as to what forms DDoS attacks against the wireless Internet will possibly take and how they can be defended effectively – technologically, economically and in terms of cost. This article tries to answer these two questions. We start by briefly reviewing the mechanism of DDoS attacks in section 2.

In section 3, we analyze new features of the wireless Internet infrastructure and possible DDoS attack forms. Since various standards for the wireless Internet are still emerging, we discuss three infrastructure schemes – the Wireless Extended Internet, the Wireless Portal Network, and the Wireless Ad Hoc Network. Intuitively, possible forms of DDoS attacks include not only ones that are found on the wired Internet – e.g., attacking e-business servers – but also new forms such as attacking the radio spectrum that is naturally a scarce resource. Another new attack form is the attack across both the wireless and wired Internet. Given the differences in computational power and the bandwidth between wired and wireless devices, it is easier for an attacker to use wired devices to initiate cross platform attacks toward wireless devices.

Section 4 proposes a conceptual model for defending against wireless DDoS attacks. In this model, we address three issues. First, we consider technological solutions based on the analysis of possible attacks. Secondly, we evaluate economic costs and benefits involved in motivating the usage of these technological solutions. As the attacks in February 2000 have shown, the biggest barrier in defending against DDoS attacks is the lack of economic incentives for Internet users to cooperate [9]. The third is the implementation issue – i.e., how to construct both technological solutions and incentive structures in a cost-effective way. Section 5 concludes this article.

2. Mechanism of DDoS attacks

The DDoS attack is the most advanced form of Denial-of-Service (DoS) attacks. As the name suggests, the DDoS attack is distinguished from other DoS attacks by its ability to deploy its weapons in a "distributed" way over the Internet and to aggregate these forces to create lethal traffic. What drives hackers to move DoS attack tools to the distributed level is the ever-increasing security in potential victims' systems in this cat-and-mouse game. Figure 1 outlines the evolution of both attacks and defenses. For a detailed explanation see [9,16].

Although the presence of bugs in network software makes the most primitive DoS attacks still viable, e-businesses are more sensitive and prompt than before in protecting their system security by using intrusion detection software and by applying patches. As a result, the most frequent and harmful DoS attacks are in distributed form. DDoS attacks are distinct from all prior DoS attacks in that they never try to break into the victim's system, thus making any security defense irrelevant. There are numerous variances of DDoS attack tools, all of which share a similar structure.

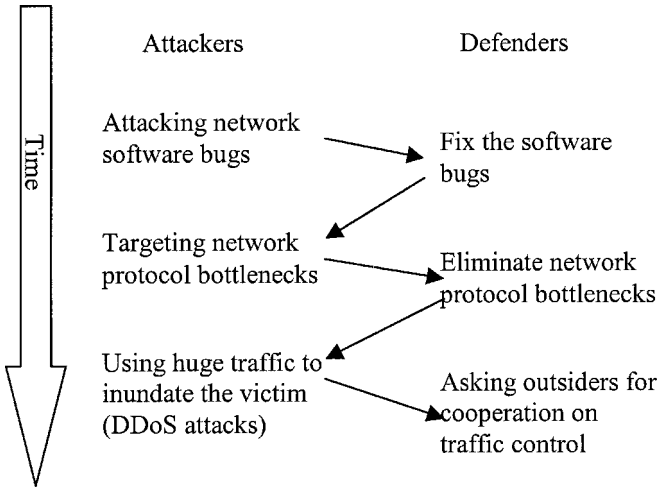

Figure 1. The evolution of attacks and defenses in DoS attacks.

A typical DDoS attack structure is shown in figure 2. The attacker first gets control of several master computers by hacking into them. Then the master computers further get control of more daemon computers (also called zombie computers), often by using some automatic intrusion software. Such a hierarchical structure is difficult to trace back. Finally, a command from the attacker can synchronize all daemons to send junk traffic to the victim, often a well-known site in e-commerce, to effectively jam its entrance and block access by legitimate users.

In practice, various DDoS tools differ in terms of the hierarchical structure, attacking packets generated, corresponding attacking targets, and the encryption of communication. For a more comprehensive list and analysis, see Packet Storm at `http://packetstorm.securify.com/` and David Dittrich's articles at `http://staff.washington.edu/dittrich/misc/ddos/`. It is worth noting that all these DDoS attack tools are available in source codes on the Internet and new versions keep emerging. New and "improved" versions are more complicated in the way they conceal attacking traffic and in encryption methods, making the defense more difficult.

For the wired Internet, Geng and Whinston [9] show that three problems lead to the proliferation of DDoS attacks: the insecure Internet, a lack of an effective way to control junk traffic, and IP spoofing.

3. Infrastructures of wireless Internet and DDoS attacks

Two aspects differentiate the wireless Internet from the wired Internet. From a technological perspective, differences between wired and wireless networks are due to link characteristics and user mobility [19]. Compared to coaxial cable, DSL, and fiber, the wireless link is characterized by high cost, volatility, high error rates and relatively small transmission capacity. Because of shared radio spectrum, communication can be interfered by competing users, other equipment, evil

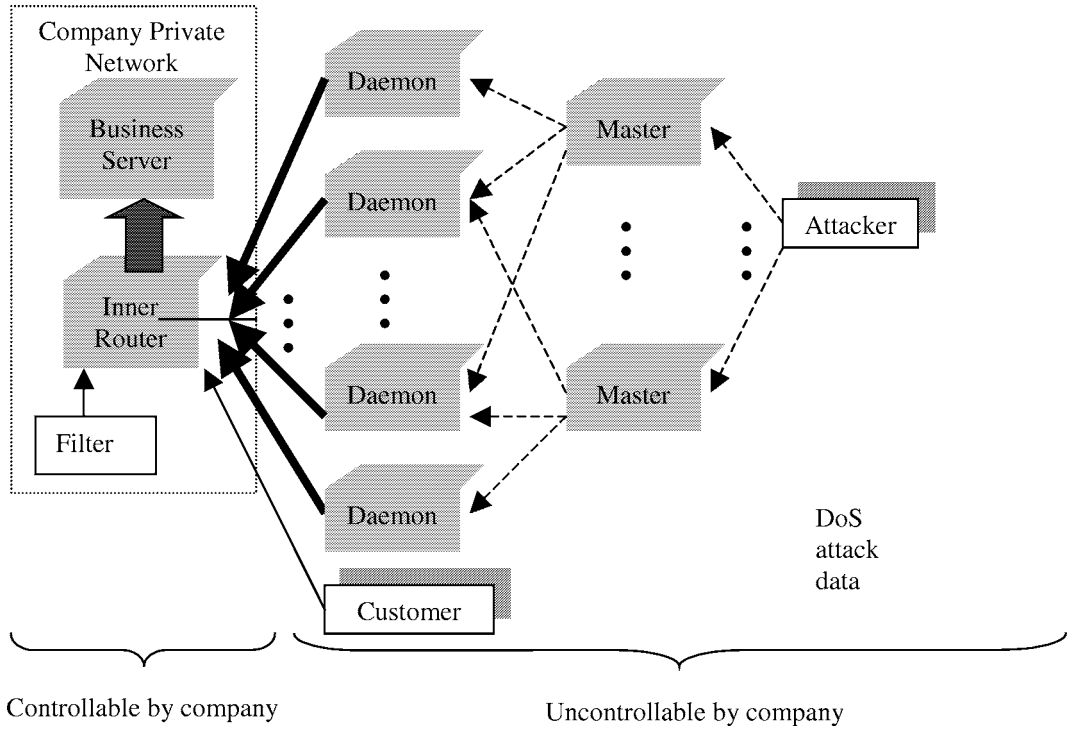

Figure 2. A typical DDoS attack structure.

intent hackers, or even natural phenomena. In terms of user mobility, the user–network interface (UNI) in a wireless environment keeps changing throughout the duration of a connection.

From an economic perspective, the wireless infrastructure is likely to be an oligopolistic market, while the wired infrastructure is open to competition. The wireless infrastructure market is dominated by a few cellular phone carriers and wireless equipment providers with different communication standards and private technologies. In addition, the high cost of radio spectrum licenses and geographic constraints make an entry to the wireless access market difficult.

Based on different application models, the wireless Internet can be categorized into three different infrastructures: the Wireless Extended Internet, the Wireless Portal Network and the Wireless Ad Hoc Network. The Wireless Extended Internet is merely an extension of the wired Internet for mobility convenience. Wireless Portal Networks are developed and privately owned by wireless telecommunication providers, thus are highly centralized. Unlike the former two, Wireless Ad Hoc Networks have no client-server structure.

3.1. Wireless Extended Internet

In the Wireless Extended Internet, wireless technology is used only for the last mile. Wireless access providers, or wireless ISPs, connect mobile devices to fixed networks via radio frequency (RF) channels. The traditional Client/Server architecture, as well as existing transport layer protocols (usually TCP), is also used for the Wireless Extended Internet. Therefore, DDoS attacks seen in the wired Internet are still feasible in the Wireless Extended Internet.

Attacking devices using aggregated traffic. Tens of millions of cellular phones, laptops and palmtops are expected to use wireless connections to access the Internet in the near future. Although transmission rates in wireless networks are much lower than those in wired networks, potential DDoS attacks are still feasible if large population of mobile units are involved. Thus, wireless data packet traffic is a potential avenue for DDoS attacks.

Attacking the asymmetric structure. Mobile devices have less computation and communication capabilities than those of fixed devices. A DDoS attack, even launched by a small number of powerful fixed computers, can effortlessly disable a large range of mobile devices. Wireless Internet content servers – such as WAP, wireless game, and mail (instant message) servers – are often optimized for small throughput and timely response. Thus, they are especially vulnerable to DDoS attacks compared with traditional wired servers.

Furthermore, there may emerge new forms of DDoS attacks taking advantage of new characteristics of the wireless communication.

Attacking the radio spectrum. The limited availability of radio spectrum is always the bottleneck in a wireless network. Even if license-free RF bands (such as the ISM band in the US) are used and micro-cell and pico-cell technologies are employed to expand transmission rates, it is still a scarce resource as the number of users and the demand for bandwidth increase. Technological research on wireless bandwidth allocation and admission control relies on stochastic theories, assuming that users will not use their devices all at the same time. Therefore, the total communication bandwidth can be

far less than the total communication capacity of all wireless devices. However, a DDoS attack deliberately coordinates wireless devices to send out synchronized traffic, which can easily consume all spectrum resources or at least significantly reduce the capacity of communication channels for normal traffic.

Avoiding tracing back by mobility. The IETF Mobile IP protocol is a significant step towards enabling nomadic Internet users. Most research on security in Mobile IP deals with registration, authentication, key management and encryption. However, Mobile IP still has flaws that DDoS attackers can use in addition to conventional security problems. For example, the Mobile IP protocol requires two IP addresses: the home address and the care-of address. The home address is permanently assigned to a mobile device, while the care-of address is temporarily assigned by the visiting foreign network. Similar to IP-spoofing, the Mobile IP protocol allows a mobile device to send out IP datagrams using its fixed home address even if it roams away. Some extensions of Mobile IP are also sources of concern. For example, the Non-Disclosure Method (NDM) prevents the tracking of user movements by third parties and gives mobile users control over the revelation of their location information, according to their personal security demands [7]. As a result, victim sites will find it difficult to trace sources of DDoS attacks.

3.2. Wireless Portal Network

Learning from America Online's success, most wireless operators are using various "walled garden" and partnership approaches. Since they own coveted spectrum licenses and cellular phone user bases, these operators have strong bargaining power over all their business partners. Therefore, they are in a better position to secure additional revenue streams, including slotting fees for portal placement, a slice of m-commerce revenues, and fees from location-based services. Such an extension of their business will transform them into wireless portals (see figure 3). The most cited example is NTT DoCoMo, for which 5.9 million users signed up with its i-mode service during the last four months of 2000.

The Wireless Portal Network is based on the typical Client/Server architecture. Mobile clients (usually cellular phones, smart phones, and specific PDAs) embedded with compact Operating Systems communicate with base stations through wireless packet-switched data networks. All requests are passed to the service center through the telephone network and signaling systems. Similar to the Service Control Point (SCP) in an Intelligent Network, the service center keeps user information and provides portal services, contracted services, and public Internet services. Portal services are kernel services in a Wireless Portal Network, which maintain user profiles and billing databases and provide location-based service and other real-time services. Application requests and responses will not be encapsulated in IP packets. Thus, they have the lowest latency. For contracted services, the requests are translated into TCP/IP protocol streams by the TCP/IP gateway and served by contracted content providers. Dedicated lines and reserved paths guarantee security and QoS. For public Internet services, Internet access requests will be passed from edge routers to the backbone.

Clients, contracted content providers, and the service center become a walled community, i.e., a reliable "security island". This architecture is more secure than the Wireless Extended Internet because a Portal Network screens all clients and most servers located in the public Internet. It is difficult to launch attacks from outside the island. However, with increasingly powerful phones, such as Java phones that could be infected with DDoS zombie viruses, the network could be vulnerable to internal attacks.

Attacking the radio spectrum. Because Wireless Portal Networks primarily employ existing cellular phone systems (single-hop), a base station is the only entry to a specified cell. In major cities and crowded airports, it is common to have calls dropped in mid-sentence. Sometimes making a connection is impossible. Mimicking this natural congestion, it is possible to disable a particular base station – e.g., the one

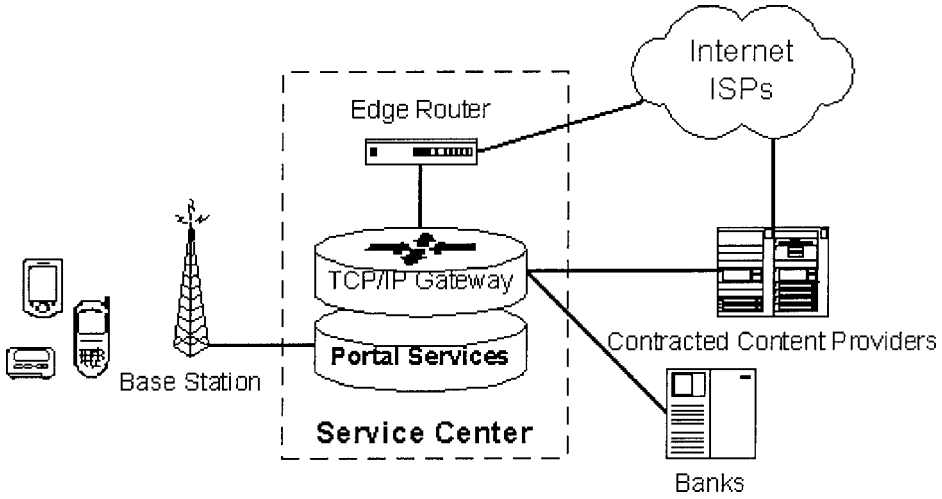

Figure 3. The architecture of the Wireless Portal Network.

serving an important conference nearby – by simultaneously sending connection requests and a mass of traffic from mobile zombies. As a result, all wireless devices within this cell will not be able to connect to the network. In some cases, even control channels can be blocked. In a Personal Communications Services (PCS) Network, when a Visitor Location Register (VLR) fails and broadcasts a re-registration request to all Mobile Stations (MSs), registration messages sent by MSs will cause a natural traffic jam (and thus, collisions) in the reverse Digital Control Channel (DCCH) [14]. Therefore, if the MSs have more control over the DCCH, they can block the channel and make VLR busy with recognizing fake identities. Then the traffic channel will be of no use even if it is available.

Attacking TCP/IP gateway. The TCP/IP gateway translates between wireless bearer protocols and the Internet TCP/IP protocols. It is one crucial bottleneck in the Wireless Portal Network. Abundant computing capability and enough links are extremely important for it to provide a security protection for mobile terminals and inner servers against attacks from the public Internet. If one has to shut down the gateway, the Wireless Portal Network will be isolated from the public Internet and make all outside services unavailable.

Attacking value-added services. It is difficult to attack value-added services since dedicated lines will be used for such crucial services as banking and trading and some content servers are embedded into portal services, like location services. All these services are invisible outside the portal networks and will survive under outside DDoS flooding. However, there might be sophisticated methods to launch attacks from devices within the portal network.

3.3. Wireless Ad Hoc Network

A Wireless Ad Hoc Network (also called multihop network or Peer-to-peer wireless network) is formed temporarily by a group of mobile devices, which have a common mission or interest. Adhering to a strict admission policy and communication rules, all these devices form a special community of equals to share information. There is no designated client or server. All members communicate over wireless channels directly without any fixed networking infrastructure or centralized administration. In this structure, all mobile hosts communicate with each other in a wireless multi-hop routing style. Each mobile node maintains all the links within the defined radius (called zone) and acts as a router in the network. If a member is out of its destination member's zone or it is not in a line-of-sight, all messages between them must pass through one or more routers. All members are free to move around and join and leave a network at will without any technical difficulties, subject to admission control. The routing scheme is adjusted dynamically according to the changing network topology.

Analogous to the Internet that evolved from the simple DARPA net, the Wireless Ad Hoc Network has the potential to grow into a World Wide Wireless interconnected network. Wireless Ad Hoc Networks were first recognized as an important issue in the military communications arena in the 70's. Several systems have been deployed for the Tactical Data Systems, such as Link-16 in the US Navy Airborne and Shipboard systems. Following the wide deployment of mature wireless technologies, the Wireless Ad Hoc Network is receiving more attention for commercial applications, such as team collaboration applications, networking intelligent sensors and cooperative robots, etc.

The Ad Hoc Network is the best architecture against DDoS attacks. First and foremost, it has no central server. Secondly, it may implement strict admission policies making it very hard for outsiders to hack into the communication infrastructure. Multi-hopping reduces transmitter power and protects network capacity via spatial reuse. Because there is no central point and no crucial resource, any blocked route can be substituted by redundant links. In addition, the community can reject an abnormal member by voting based on certain admission policies. Dynamic routing protocols and mobility of the network components give Ad Hoc Networks a self-adjusting capability under attacks.

It is unlikely that the Wireless Ad Hoc Network will be restricted to a small geographical region. Hybrid architecture could be used to expand the range of such networks. Members can communicate with one another via the local RF network within a regional wireless community, and with other members located anywhere within reach of the commercial telephone system through wired relay services. With the help of the dual-membership hosts, interconnecting different communities will result in the World Wide Wireless network. Wireless communities can also be attached to conventional fixed data networks to expand application possibilities. For instance, home-networked appliances based on Bluetooth technology can be remotely controlled through the Internet. For military use, a complete networking system, called the AEGIS Broadcast Network, has been implemented for tactical data systems in the US Navy. It connects, monitors, and controls all military units on both coasts, the Gulf of Mexico, Japan, etc. The interconnection among Wireless Ad Hoc Networks through wired relay services creates a complex network topology, in which critical points can be attacked. First, attacks against dual-membership hosts may effectively disable the interconnections among different Ad Hoc Networks. Secondly, directory services, which are indispensable for large scale interconnected Ad Hoc Networks, are also possible targets for DDoS attacks. This is similar to the case in the Internet where DNS servers and catalog servers are frequent targets of DDoS attacks. In a word, the World Wide Wireless network could be subject to all forms of DDoS attacks that exist on the Internet if it evolves towards an asymmetric infrastructure.

4. Defending against DDoS attacks on the wireless Internet

In the event of a typical DDoS attack, the victim alone cannot effectively defend herself/himself. Cooperation among all involved parties is indispensable. Figure 4 presents our conceptual model for defending against a DDoS attack, which illustrates a two-layer coordinated defense problem and an implementation problem.

In the two-layer coordinated defense problem, the first layer focuses on effective coordinated technological solutions. The second layer deals with the incentive mechanism that, in an economic perspective, makes people involved in a DDoS attack feel that cooperating with each other is the best strategy. In past practice, unfortunately, little attention has been paid to this second layer problem compared with the public focus on technologies. Ironically, this incentive problem causes the most headaches in practice [9]. As a solution, we propose to use usage-based fees as the foundation of the incentive mechanism.

The objective of the implementation problem is cost-effectiveness, which arises as a crucial problem because defending against DDoS attacks may require an overhaul of the current network infrastructure. For instance, the implementation of a usage-based fee scheme on the wireless Internet – as well as on the wired Internet if we consider the cross-border attacks between the wired and wireless Internet – has strong demands on the network's ability to audit and manage traffic. As an illustrative example, we present an implementation scheme based on the Policy Based Networking (PBN) framework.

4.1. Coordinated technological solutions

There are four types of coordinated technological solutions, as shown in figure 5 [9].

Two comments are necessary for figure 5. First, different solutions can coexist to achieve a better defense. For example, user-level traffic control and coordinated filters can be implemented simultaneously to be more effective. Second, as in the wired Internet example, coordination is often required to be global, whereas in the wireless Internet case local coordination may suffice. For example, to avoid an attack on radio frequencies in a certain geographical area, it is sufficient to require coordination only among involved wireless devices and base stations in that area. Below we analyze the characteristics of these four coordinated technological solutions.

Improving the security of all relevant devices. Before initiating an effective DDoS attack, the attacker needs to break into enough zombie devices to secure an ability to generate sufficient traffic. A direct counterstrike is to secure all devices to make it difficult for the attacker to seize enough zombies.

It is not practical, nor potentially beneficial, to secure all computers on the wired Internet [9]. Alternatively, an effective and efficient solution would be to selectively secure those computers that have high traffic throughput – such as routers –

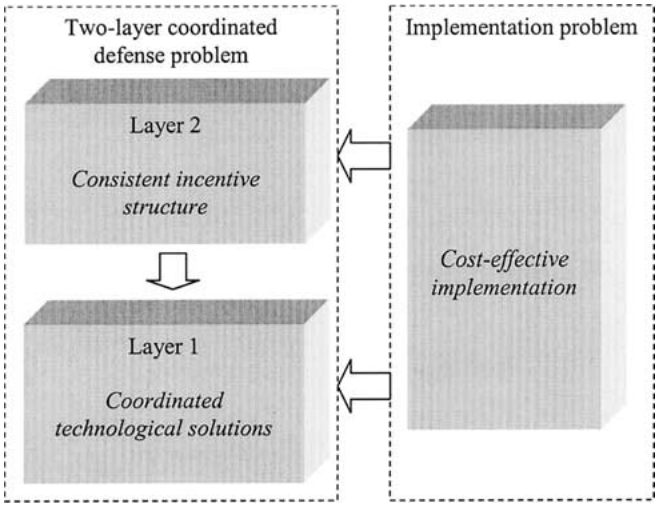

Figure 4. The conceptual model for defending against the DDoS attack.

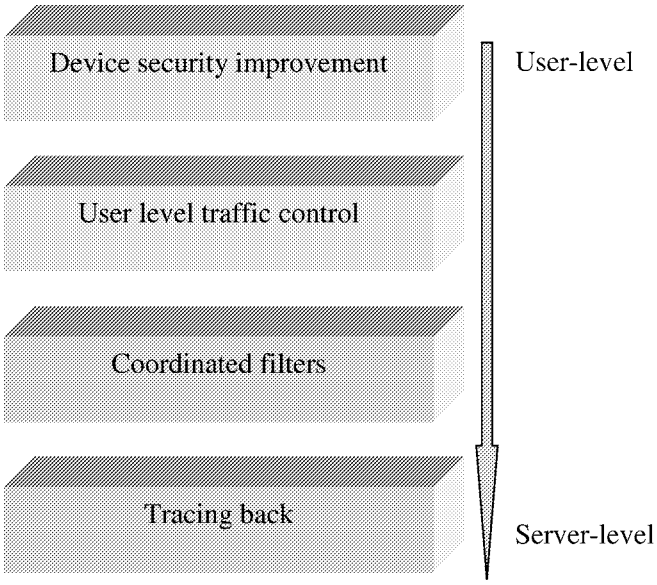

Figure 5. Four coordinated technological solutions to DDoS attacks.

or high performance and high bandwidth workstations so that the marginal benefit for each dollar spent on security is optimized. Moreover, for some networks that have the ability to audit real-time traffic, security measures can even be delayed until a DDoS attack actually happens, thus making them more targeted and therefore more efficient.

For the wireless Internet, such a selective security implies that wireless devices with high bandwidth connections, e.g., 3G devices, are the ones that should be safeguarded. We note that the wireless communication industry has a tighter security tradition than the wired Internet community, partially because of the relatively large communication spectrum and device costs.

User-level traffic control. User-level traffic control is embodied in a set of traffic control rules specifically for a given network device. For example, a wireless device user can set up a daily traffic cap that is high enough not to disturb her/his

normal usage, while abnormally large traffic will be stopped. Furthermore, the abnormal traffic may trigger a warning to the user or to a network administrator for follow-up diagnosis. Traffic control rules can be contingent on factors including other users' usage status. For example, a user can specify her/his data to be dropped or delayed if the network is experiencing congestion.

Geng and Whinston [9] propose to use an e-stamp model to control traffic even if user devices are hacked. A direct implication is that user-level traffic control rules for a specific network device need to be protected more securely than the network device itself since we do not want the attacker to modify the traffic control rules once she/he gets control of a network device. For the wired Internet, Geng and Whinston propose to save the rules in edge routers because routers, given their concentrated and limited functionalities, are relatively easier to protect than other computers.

For the wireless Internet, the candidate host for traffic control rules can be flexible. Unlike desktop computers that are normally anonymous with concealed identity, wireless devices – especially wireless phones – have unique IDs or PINs that are transmitted along with the data which cannot be tampered with. These IDs or PINs can be used to identify wireless devices. Furthermore, unlike desktop computers in which software programs can control and modify virtually all information including the traffic control information, wireless devices normally have restricted access functions that enable secure traffic control even if the wireless device is hacked.

Edge routers in the Wireless Extended Internet and gateways in the Wireless Portal Network are the ideal hosts for coordinating user-level traffic control rules. For example, if a user wants her/his data packets to be dropped when the outbound network of the wireless ISP is congested, the edge router has the ability to realize this requirement. The designation of a host for traffic control rule coordination is complicated in a Wireless Ad Hoc Network since no one party is more likely to be in a central position than another.

Coordinated filters and tracing back. Even when user-level traffic control fails, wireless ISPs in the Wireless Extended Internet can still try to defeat DDoS attacks by identifying the attacking traffics and stopping them by using coordinated filters. The purpose of coordination among filters is to stop the traffic as early as possible along the attacking paths to prevent the damage from aggregated traffic. In a Wireless Portal Network, due to the relatively simple network topology, coordinated filters can be simplified to only one single filter. For a Wireless Ad Hoc Network, filtering is not applicable due to the symmetric structure. However, community rules, e.g., a voting mechanism, may play the role of a central filter to decide which user device to block.

Even if the coordinated filters cannot effectively stop the attack, possibly because the attacking traffic is hard to distinguish from normal traffic, there still exists another technological solution – to trace back to the zombie devices (and possibly the attacker) to shut down the attack from the source. Combining this with possible legal actions, this method can also help to deter repeated attacks.

4.2. A consistent incentive structure

According to the Yankee Group, a Boston consulting firm, the DDoS attack in February 2000 cost approximately $1.2 billion, not to mention the damage to consumer confidence in e-commerce [18]. Effective coordinated solutions to DDoS attacks are critical for the future of e-commerce and m-commerce. However, a fervent advocacy of coordinated solutions does not necessarily result in actual implementation. Sample research by icsa.net, for example, shows that less than 15 percent of all corporate users are filtering source IP addresses. An even smaller percentage of Internet service providers – less than 8 percent – are doing this type of filtering [15].

A disincentive structure for the wired internet. The reason for this low rate of implementation of coordinated solutions is the inconsistent incentive structure in Internet traffic pricing. Simply stated, the victim has the incentive to defend but cannot defend effectively, whereas the owners of zombie computers and ISPs can defend effectively but do not have the incentive to do so. In this time of flat monthly fee payments for wired Internet access, the owner of a zombie computer incurs little cost due to DDoS attacks since all that is stolen is just some traffic. On the other hand, preventing a personal computer from being controlled by any potential attacker requires frequent – virtually constant – monitoring and updating, at considerable cost. If the cost of protection is higher than the value of the traffic being protected, an economic disincentive clearly exists. Similar logic applies to ISPs who can always collect the monthly fees no matter whether a DDoS attack happens or not. Thus, they may hesitate to install filters since they will lower network performance.

Who should be motivated to defend in the wireless Internet? Having observed the failed incentive structure of the wired Internet, it is clear that the wireless infrastructure should contain a new incentive structure that can give wireless device owners and ISPs enough impetus to implement defense mechanisms. However, an efficient incentive structure need not target all wireless device owners – only high-bandwidth devices should be effectively protected, including:

- high-performance, high-bandwidth end-user devices (including wired devices that can communicate with the wireless Internet),
- routers, and backbone switches.

As we mentioned before the possibility of attacking the Wireless Extended Internet from the wired counterpart, the incentive structure is also need for devices in the wired Internet. As wired devices generally have more communication capacity, the incentive structure for the wired network needs to be more strict.

Table 1
i-mode pricing scheme (US $1 = Japan ¥123.5 as of July 12, 2001). According to MobileInfo.com (http://www.mobileinfo.com/imode/buz_approach.htm), the average monthly bill is $30–$40 (or ¥3700–¥4900). Therefore, part 2 is the leading cost.

Part 1: monthly charges	Part 2: packet transmission charges	Part 3: i-mode information charges
¥300/month	¥0.3 per packet (128 bytes)	Usually ¥100–¥300/month for each fee-based service

An incentive structure based on usage-based fees. One candidate for an effective incentive structure is the usage-based fee. The direct effect of a usage-based fee is a sharp increase in the cost to zombie devices if they are sending out attacking traffic. In particular, if a proper fee increase scheme is devised, it should not affect normal network usage but the cost could increase significantly for high-performance, high-bandwidth devices when they are sending out huge traffic volume.

These computers are most often located in corporations, governments, and universities. With a usage-based fee structure, the owners of such computers will have the greatest immediate incentive to take security actions. Similarly in the wireless Internet, devices that have the potential to occupy a large portion of the radio frequency will be controlled most tightly. Likewise, a usage-based fee between an ISP and a backbone provider encourages the ISP to have more concern over its traffic. Specifically, such a usage-based fee plan makes ISPs more likely to install coordinated filters and to support user-level traffic controls.

Fortunately and unlike the wired Internet industry, the wireless Internet industry starts with usage-based fees. For example, Japanese vendor DoKoMo's i-mode service pricing is mainly packet based, as shown in table 1.

US wireless providers are using minute-based pricing plans that are often simplified (as we will explain shortly) to the form of fixed pricing with an over-the-cap penalty for several service levels. Currently given the low bandwidth and simple functions of wireless devices in the US, simple pricing schemes based on connection time are applicable. However, it is conceivable that with the increase of bandwidth and more rich applications with different traffic requirements, and more importantly with the migration to packet-based communication, packet-based pricing will become more accurate and practical than minute-based pricing.

The wireless Internet: Towards dynamic usage-based fees. If the usage-based fee continues in the wireless Internet, we can expect less DDoS attacks compared with the wired Internet. A usage-based fee can be further calibrated to provide more targeted incentives against DDoS attacks, i.e., a dynamic usage-based fee plan can better prevent DDoS attacks than constant usage-based fees [3,11]. A constant usage-based fee scheme has a fixed unit price. Packet-based pricing is an example of the constant usage-based fee, while the dynamic usage-based fee implies a changing unit price, which is higher when there is congestion in the network [3,11].

Wireless service providers (as well as long-distance phone providers) have already considered predictable congestion for their constant usage-based fee scheme. For example, it is a common practice to price higher for daytime communication than for nighttime or weekend communication as congestion is more likely to happen in daytime. We call this the modified constant usage-based fee scheme.

A dynamic usage-based fee scheme, on the other hand, deals with unpredictable congestions, including those caused by DDoS attacks. The characteristic of a dynamic usage-based fee is the increase in unit price when congestion happens or will happen. The incentive it gives to wireless device owners is twofold. First, those owners are more likely to set up traffic control rules in their device to instruct to delay or cancel the data transmission when the network is congested or approaching congestion. Therefore, even if an attacker instruct all zombie devices to send attacking traffic at the same time, an effectively synchronized attack is unlikely to occur. Second, as congestion means higher cost, high bandwidth owners are more likely to invest more in the security of their devices to avoid stolen traffic.

Table 2 gives a concise comparison of three usage-based fee schemes.

Usage-based fees can be flexible. It is constantly questioned whether or not users will accept a usage-based fee plan even when it is financially beneficial for them. Some researches [20] show that many people dislike the uncertainty and complexity associated with usage-based fees. Concerning this problem, it is worth pointing out that a consistent incentive structure can be flexible in its form while still representing the essence of a usage-based fee plan, as illustrated in table 3.

For the Wireless Ad Hoc Network, a monetary incentive structure may not be available simply because of the lack of a charging system. Instead, other incentive mechanisms, e.g., a voting mechanism which effectively rules out a member upon heavy radio frequency usage, can serve the same purpose.

Once again, for defending the Wireless Extended Internet, a usage-based fee plan is also needed for the wired Internet. Nevertheless a usage-based fee plan for the wired Internet is mainly used to prevent DDoS attacks inside the wired Internet, for which Geng and Whinston [9] have discussed possible mechanisms.

4.3. Cost-effectiveness

The history of the Internet shows that the de facto criteria for success in any proposal are whether that solution is proactive and consistent with mainstream and commercial Internet technologies. Because of the anonymous and "best effort" usage of the Internet, it is arduous and costly to regulate the infrastructure against DDoS attacks. Several advanced network management technologies have been proposed to address the traffic control problem. Employing these existing technologies will significantly reduce the costs and risks in designing future wireless Internet.

Table 2
Different usage-based fee schemes.

Fee scheme	Characteristics	Effects	Examples
Constant usage-based fee	Fixed unit (unit traffic volume or unit communication time) price	Provides basic incentive for wireless device users to prevent traffic stolen	i-mode packet-based pricing
Modified constant usage-based fee	Multiple fixed unit prices, each for a given time period	Prevents/alleviates predictable congestions	Price differentiation for daytime and nighttime (e.g., SprintPCS)
Dynamic usage-based fee	Unit price increases when congestion happens or will happen	Prevents/alleviates any possible congestions	N/A

Table 3
Variations of constant usage-based fees.

Fee scheme	Examples	Comments
Constant usage-based fee	Minute-based fee (many long-distance services in US), packet-based fee (i-mode)	The most preliminary usage-based fee scheme. Users are exposed to financial risks as they may receive large bills.
Constant usage-based fee with a cap	Pre-paid phone cards	An upper cap prevents financial risks for users. Nevertheless, the cap may be reached when important communication is going on.
Flat monthly fee with a communication cap	The next scheme is similar to this one. To our knowledge, now most plans allow over-the-cap usage for a higher price.	Limited communication volume/time implies that this is still usage-based pricing. Often several plans of different caps are offered to let users to self-select. Reduces the complexity of usage-base fee.
Flat monthly fee with over-the-cap penalty	Most cell phone plans from most US providers	High over-the-cap penalty (often around \$0.5/min) effectively stimulates usage control.

The Policy Based Networking (PBN) [24] is one promising technology for implementing usage-based fees to deal with DDoS attacks. Essentially, it provides rules that describe actions to take when specific conditions arise. These policies are able to control critical network resources such as bandwidth, QoS, security and Web access across heterogeneous networks. Thus, both natural and artificial congestions are under the control of a globally coordinated structure. As illustrated in figure 6, we present an implementation scheme based on the PBN, and discuss how to incorporate both the incentive structure and the technological solutions into this scheme in a cost-effective manner.

In this scheme, the two main elements for policy control are the Policy Enforcement Point (PEP) and the Policy Decision Point (PDP) [24]. From the PBN perspective, the Wireless Location Register/Authentication Center is a natural policy server (i.e., PDP) with additional functionality such as user authentication, accounting, and policy information storage. At network border points, PEPs act as a "police" to accept or deny requests appropriately. Through secure and reliable channels (such as telecommunication out-of-band signaling network), PDPs and PEPs can exchange policy information with the Common Open Policy Service protocol (COPS) [2].

At the user's end, with the Intelligent IC card and other hardware technologies, wireless devices have some embedded functionalities that cannot be tampered with. The user-end policies have three levels. First, providers can deploy policies in terminals which users cannot change. Unlike desktop computers that are normally anonymous in the sense that they can conceal their identities, wireless devices such as wireless phones have unique IDs, or PINs, that are transmitted along with the data and cannot be altered. These IDs or PINs are effective instruments to identify wireless devices. Also, there are restricted access functions, such as integrating admission control into lower layer traffic control to increase the performance and security [5]. These restrictions can enable secure traffic control of all relevant devices even if these devices are hacked.

Second, end users could design their own policies, which are unchangeable by applications. For example, a user can assign a daily cap in traffic for her/his cellular phone. If the cap is reached, the system could block any further transaction and/or raise an alarm. In fact, the pre-paid cellular phone card implements a similar traffic-cap function. Future mobile phone users can set rules that are more sophisticated.

The above two policy controls cannot be realized without specific hardware that is configurable only by providers or end users. A third level policy control can be constructed in software by enabling a wireless operating system to have multiple security levels. Policy control is realized in higher security levels that normal networking applications cannot modify.

Finally, at the Intranet border point, TCP/IP gateways play the role of policy proxies. Proper policy rules can turn these proxies into coordinated filters and even support advanced usage-based fee schemes, such as dynamic pricing. The entities involved in policy control can verify each other's identity and establish necessary trust links before communicating. With the help of standard PEPs on Internet edge routers,

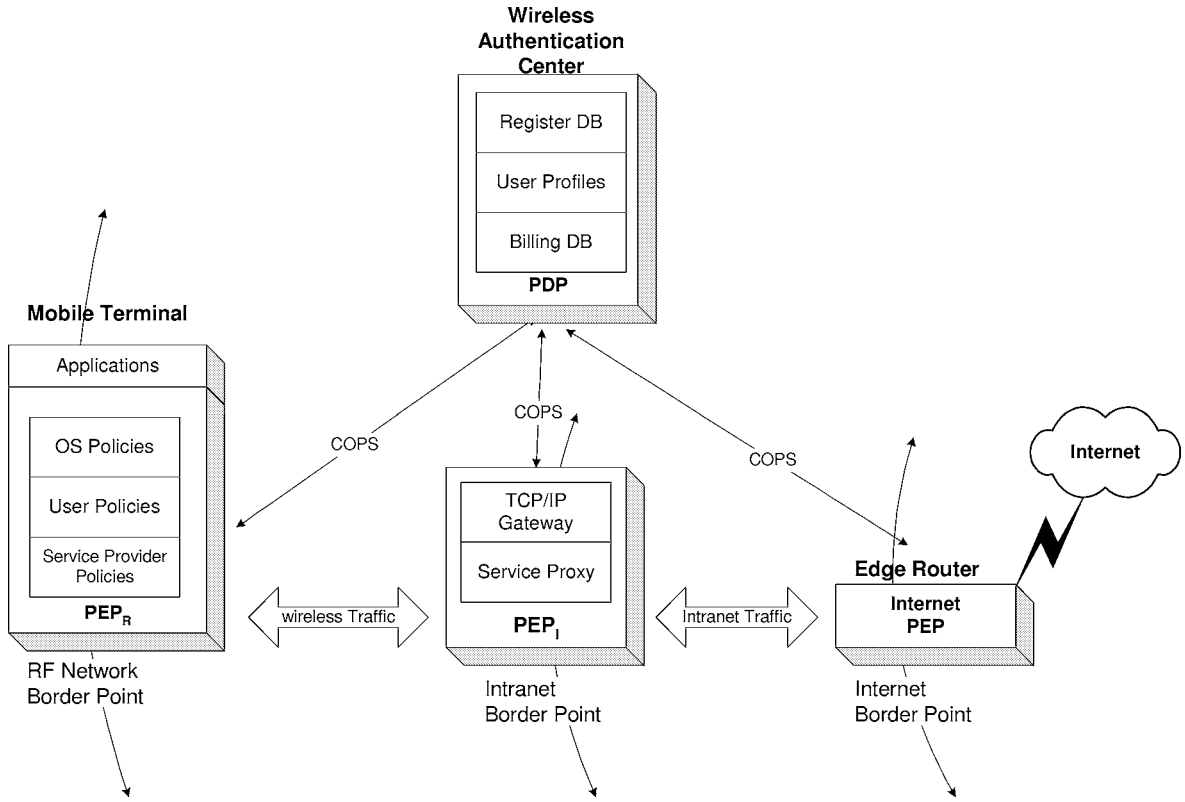

Figure 6. A wireless network architecture based on the PBN.

a global coordinated network will be formed to minimize theft and DDoS threats.

A usage-based fee scheme can be implemented by using PDPs and PEPs, for example, in the following way. First, once the fee scheme is decided, it is implemented as a set of policies in PDPs at the Wireless Authentication Centers. Secondly based on the fee scheme and the real-time traffic condition, a PDP decides the pricing rules for every related mobile terminal and send these rules as policies to PEPs on these mobile terminals. Thirdly PEPs on mobile terminals enforce these pricing rules. Whenever there is a surge in traffic, possibly caused by DDoS attacks, PEPs report the traffic change and any possible congestion to the coordinating PDP, who in return dynamically adjusts pricing rules according to the given fee scheme and instructs PEPs to update their pricing rules.

5. Concluding remarks

The DDoS attack threatens all time-sensitive m-commerce services. Fortunately the wireless Internet currently has a distinctive advantage over the wired Internet in defending against the DDoS attack: the timing. When DDoS attacks came to the wired Internet, the infrastructure of the wired Internet had been stable for decades, albeit lacking reliable mechanisms for QoS control and incentive structures for traffic control. As a result, it was repeatedly targeted by DDoS attacks. In comparison, the wireless Internet industry has a chance to address DDoS attacks before it fully matures. However, time is running short as a well-founded wireless Internet infrastructure is expected to emerge by 2003 [10]. Whether potential DDoS attacks on the wireless Internet will materialize or not will solely depend on how the wireless industry deals with the potential problem when solutions can still be embedded into the basic infrastructure.

References

[1] Bear Stearns, Mobile Internet and applications, Research report (June 2001).

[2] J. Boyle, R. Cohen, D. Durham, S. Herzog, R. Raja and A. Sastry, The COPS (Common Open Policy Service) Protocol, IETF RFC 2748, Proposed standard (January 2000).

[3] CREC, Congestion pricing more profitable than AOL pricing, Research report, http://crec.bus.utexas.edu/news/nr.html

[4] CTIA, Background on CTIA's semi-annual wireless Industry survey, http://www.wow-com.com/pdf/wireless_survey_2000.pdf

[5] S.K. Das, R. Jayaram, N.K. Kakani and S.K. Sen, A call admission and control scheme for quality-of-service (QoS) provisioning in next generation wireless networks, Wireless Networks 6 (2000) 17–30.

[6] S. Dennis, Mobile phones emerge as new virus target Kaspersky, Newsbytes.com, http://www.newsbytes.com/news/00/153195.html

[7] A. Fasbender, D. Kesdogan and O. Kubitz, Analysis of security and privacy in Mobile IP, in: *4th International Conference on Telecommunication Systems, Modeling and Analysis* (1996).

[8] D. Fisher and D. Callaghan, Microsoft attack raises concern over new DDoS variant, Yahoo! News (January 26, 2001) http://dailynews.yahoo.com/h/zd/20010126/tc/microsoft_attack_raises_concern_over_new_ddos_variant_1.html

[9] X. Geng and A.B. Whinston, Defeating distributed denial-of-service attacks, IEEE IT Professional (July/August 2000) 36–41.
[10] Goldman Sachs, Technology: Mobile Internet, Research report (September 2000).
[11] A. Gupta, D.O. Stahl and A.B. Whinston, The economics of network management, Communications of the ACM 42(9) (September 1999) 57–63.
[12] C. Haney, Network Associates hit by denial-of-service attack, IDG News Service (February 2, 2001) http://www.computerworld.com/cwi/stories/0,1199,NAV47-68-84-88-93_STO57290,00.html
[13] D.I. Hopper, Denial of service hackers take on new targets, CNN.com (February 9, 2000) http://www.cnn.com/2000/TECH/computing/02/09/denial.of.service/
[14] Z.J. Haas and Y. Lin, Demand re-registration for PCS database restoration, Mobile Networks and Applications 5 (2000) 191–198.
[15] ICSA.Net, 650-member Alliance for Internet Security unveils tool to detect network vulnerability, http://www.icsa.net/html/press_related/2000/3_23_00_NetLitmus.shtml
[16] Internet Security Systems, Denial of Service FAQ, http://www.iss.net/news/denialfaq.php
[17] Kaspersky Labs Int. demystifies the discovery of the first "true" wireless virus, http://www.avp.ru/news.asp?tnews=0&nview=1&id=107&page=0
[18] D. Murphy, Recent "Denial of Service" attacks cost $1.2 billion, InsiderReports.com, http://www.insiderreports.com/storypage.asp_Q_ChanID_E_WB_A_StoryID_E_20000526
[19] M. Naghshineh, M. Schwartz and A.S. Acampora, Issues in wireless access broadband networks, in: *Wireless Information Network*, ed. J.M. Holtzman (Kluwer Academic, 1996).
[20] A. Odlyzko, Internet pricing and the history of communications, Draft, http://www.research.att.com/~amo/doc/networks.html
[21] L. Sherriff, Virus launches DDoS for mobile phones, http://www.theregister.co.uk/content/1/12394.html
[22] Strategy Analytics, Strategy Analytics forecast $700B wireless market by 2005, http://www.strategyanalytics.com/press/PRDK007.htm
[23] H.R. Varian, Economic scene: Liability for Net vandalism should rest with those that can best manage the risk, the New York Times (June 1, 2000).
[24] R. Yavatkar et al., A framework for policy based admission control, IETF RFC 2753 (January 2000).

Xianjun Geng is a doctoral student in the Department of Management Science and Information Systems at the Graduate School of Business, University of Texas at Austin, and a research associate at the Center for Research in Electronic Commerce. His research focuses on the impact of emerging technologies on the digital economy. He is also interested in the use of economic principles to analyze competition and marketing issues in electronic commerce.
E-mail: gengxj@mail.utexas.edu

Yun Huang obtained his degrees Bachelor of Science and Master of Science (both in computer science), respectively, in 1997 and 1999, all at the Tsinghua University, in P.R. China. He is currently a Ph.D. candidate in the Department of Management Science and Information Systems and a research associate at the Center for Research in Electronic Commerce, both at the University of Texas at Austin. His current research interests cover mobile commerce, software industry, and Internet security and competition.
E-mail: yun@mail.utexas.edu

Andrew B. Whinston holds the Hugh Cullen chair of information systems, computer science, and economics at the University of Texas at Austin. He obtained his Ph.D. from Carnegie-Mellon University in 1962 and subsequently held academic posts at the University of Texas at Austin, Purdue University, the University of Virginia, Yale University, and the University of California, Los Angeles. He was awarded the Ford Foundation Faculty Research Fellowship in 1966. Dr. Whinston's research has explored artificial intelligence, e-commerce, information systems, and the New Economy. He is Editor-in-Chief of Decision Support Systems, a member of the editorial board of Annals of Mathematics and Artificial Intelligence, and Editor-in-Chief of the Journal of Organizational Computing and Electronic Commerce. He is also the Director of the Center for Research in Electronic Commerce.
E-mail: abw@uts.cc.utexas.edu
WWW: http://crec.bus.utexas.edu