

Statisztikus fizikai módszerek az Internet forgalom aktív méréseiben

Doktori értekezés tézisei

Hága Péter

ELTE TTK Fizika Doktori Iskola

Statisztikus fizika, biológiai fizika és kvantumrendszerek fizikája program

Iskolavezető: Dr. Horváth Zalán, az MTA rendes tagja

Programvezető: Dr. Vicsek Tamás, az MTA rendes tagja

Témavezetők: Dr. Csabai István, PhD, habil. docens és

Dr. Vattay Gábor, egyetemi tanár

ELTE TTK Komplex Rendszerek Fizikája Tanszék



2007.

Bevezetés

Az emberek számára az „Internet” szó általában az e-mail és web hozzáférést, közösségi oldalakat, on-line vásárlást, vagy ügyintézését jelent. Azonban az Internet sokkal több ennél. Az Internet egy komplex felépítésű, nagy számú számítógépet és egyéb hálózati eszközöket tartalmazó, hálózatba szerveződött rendszer. A hálózat működését nagyszámú mikroszkopikus jelenség együttese határozza meg. Az Internetet, mint gráfot vizsgálva ezek az elemi jelenségek lehetnek például az új élek és csúcsok megjelenése, vagy azok eltűnése. Az Internet működését, ha azt egy kommunikációs rendszernek tekintjük, olyan elemi jelenségek határozzák meg, mint az egyes folyamatok és csomagok egymással és a fizikai rendszerrel való kölcsönhatása. A teljes rendszert rengeteg ilyen kis jelenség hajtja előre, így végül a rendszer egészének tulajdonságaira az egyedi jelenségek együttesének vizsgálatával tudunk következtetni.

Ezeknek a tulajdonságoknak a következtében az utóbbi időkben megnövekedett a statisztikus fizikai megközelítést alkalmazók száma, hiszen a rendszer megértése megköveteli ezt a gondolkodásmódot. A statisztikus fizika olyan klasszikus eszközeit alkalmazva, mint a diffúziós folyamatok, sejt-automaták, átlagtér módszerek, vagy akár a perkolációs elmélet, már számos nagyszerű eredmény született mind a hálózati topológia, mind a forgalom modellek területén.

A számos érdekes terület közül a dolgozatban a hálózati forgalomnak, mint sztochasztikus folyamatnak a vizsgálatával foglalkoztunk, hiszen a hálózati forgalom rendkívüli mértékű növekedése megköveteli, hogy a forgalom tulajdonságait, statisztikus jellemzőit, vagy akár dinamikus jelenségeit leíró pontos modellekkel rendelkezünk. A számítógépek kommunikációjának terén számos alapvető fontosságú kérdést lehet megfogalmazni, amelyek közül az utóbbi évek egyik legérdekesebb kérdése, hogy a hálózat két végpontja között mekkora pillanatnyi sebességgel lehet adatot forgalmazni. A válasz megfogalmazásához a két végpont közötti útvonal jellemzőinek ismerete szükséges. Ezen jellemzők közül a legfontosabbak a végpontokat összekötő hálózati elemek fizikai kapacitása, azaz sáv szélessége, valamint ezen eszközök időben változó szabad kapacitása, az ún. rendelkezésre álló sáv szélesség. Azonban az óriási számú egymástól független alkotóból álló rendszer legtöbb jellemzője nem tanulmányozható közvetlenül, csak közvetett módon szerezhetünk tudomást számos érdekes és fontos jelenségről, azok jellemzőiről. A hálózati forgalom tulajdonságainak meghatározására használt mérési módszereket két nagy csoportba oszthatjuk, ezek az ún. aktív és az ún. passzív mérési módszerek. A passzív módszerek alkalmazása során egy adott hálózati eszközben, lokálisan gyűjtött adatok alapján határozhatunk meg különféle mennyiségeket. Ezek a módszerek azonban csak akkor alkalmazhatóak, amennyiben hozzáférünk a kérdéses hálózati eszközhöz.

Az aktív módszerek általában valamely útvonal tulajdonságait igyekeznek meghatározni, és nem szükséges az útvonalat alkotó hálózati elemekhez közvetlen hozzáféréssel rendelkezni. Az aktív módszerek alapját ún. próbacsomagok adott

mintázat szerinti indítása jelenti, amely csomagok kölcsönhatnak a hálózat eszközeivel, illetve a hálózati háttérforgalommal. Ugyan a mikroszkopikus részletek meglehetősen különbözőek, az aktív méréseket a fizikai szórás kísérletek analógiájának is tekinthetjük. A mérésben használt próbacsomagok a hálózati háttérforgalom mértékétől függően különböző mértékű várakozásokra, sorbanállásokra kényszerülnek. A próba forgalom és a hálózat többi része közötti kölcsönhatásnak az eredményeként megváltozik a próbacsomagok mintázata. A csomagok közötti távolság megváltozásának, azaz a mért diszperziós görbéknek az értelmezése általában nem egyszerű feladat, hiszen a rendszer paramétereiről általában csak közvetett információt tartalmaznak.

Alkalmazott módszerek

A dolgozatban a csomagpár és csomagvonal diszperziós görbére adtunk különböző közelítéseket. Elsőként egyszerű megfontolások alapján egy empirikus modellt állítottunk fel, ami lehetővé teszi releváns hálózati paraméterek meghatározását. A rendszer pontos folyamatainak megértéséhez azonban részletesebb modellezés szükséges. Ehhez transziens sorbanállási elméletek segítségével adtuk meg a diszperziós görbe alakját.

A csomagvonal diszperzióját a rendszer torlódott és nem torlódott fázisai közötti kritikus pontban vizsgáltuk. A fázisátalakulások elméletéből ismert véges méret skálázással ebben a pontban meghatároztuk a háttérforgalom érkezési folyamatát jellemző exponenst.

A bemutatásra került analitikus eredmények ellenőrzésére és a bevezetett mérési módszerek vizsgálataihoz három különböző általunk létrehozott kísérleti rendszert használtunk. Az ideális kísérleti környezetek előállításához, valamint a széles paraméter tartományokban történő vizsgálatokhoz szimulációs környezetet, az analitikus eredmények ellenőrzött körülmények közötti, valódi rendszerekben való teszteléséhez laboratóriumi teszhálózatokat használtunk. Végül nagy skálás hálózati kísérleteket végeztünk nem kontrollált körülmények között az Etoic Hálózati Mérési Infrastruktúra Európa szerte telepített mérőpontjai között. A kísérletek során az általunk kifejlesztett szoftverek segítségével gyűjtöttük és értékeltük ki az adatokat.

Tézisek

1. Megmutattuk, hogy az általánosan használt ún. folyadék-modell paramétere nem elégségesek a kísérletileg megfigyelhető diszperziós görbék pontos leírásához, ahhoz a hálózati forgalom granuláris természetét is figyelembe kell venni. Egy adott útvonal fizikai sávszélességére, illetve a rendelkezésre álló

sávszélességének meghatározására adtunk egy empirikus alapokon nyugvó formulát, ami jól illeszkedik a megfigyelhető diszperziós görbékhez azok teljes tartományán.

2. A kísérletileg megfigyelhető diszperziós görbék pontos leírásához egy sztochasztikus modellt készítettünk. Ebben a modellben a háttérforgalmat egy $M/G/1$ sorban állási modellel közelítettük, amelyben a háttérforgalmat tetszőleges méreteloszlású és Poisson érkezési statisztikájú csomagok alkotják. A sorhossz eloszlásának időfejlődését a Takács-féle integro-differenciál egyenlet adja meg, ebből kiindulva határoztuk meg a csomagpár diszperziót. Zárt alakban megadtuk a csomagpár diszperzió értékét a csomagok indítási távolságának, valamint a modellben szereplő fizikai sávszélességnek, a háttérforgalom átlagának és a háttérforgalom csomagméret eloszlásának függvényében. Az így kapott diszperziós görbe helyesen írja le a kísérletileg megfigyelhető görbéket, rámutatva, hogy a megfigyelhető diszperziós görbe pontos leírásához nem hanyagolhatjuk el a háttérforgalom csomagok véges méretét.
3. Bevezettük a granuláris modellt, amely egy $M/D/1$ sorban állási modell, amelyben a háttérforgalmat Poisson folyamat szerint érkező, egyforma méretű csomagok alkotják. Megmutattuk, hogy egy effektív csomagmérettel, az ún. granularitás paraméterrel helyettesíthető bármely tetszőleges háttérforgalom csomagméret eloszlása. Így egy általános csomagméret eloszlás mellett az előző tézispontban bemutatott $M/G/1$ modellel felvett diszperziós görbe, és a granularitás értékének megfelelő, konstans méretű csomagok mellett az $M/D/1$ modellel felvett diszperziós görbe alakja közel azonos. Tehát a granuláris modell közelítésben a diszperziós görbe leírásához a folyadék modellben is meglévő fizikai sávszélességen és háttérforgalom átlagán túl a granularitás értékének megadása már elégséges, nem szükséges a teljes csomagméret eloszlás ismerete.
4. Megalkottunk egy új mérési és kiértékelési módszert, amely lehetővé teszi, hogy a hálózati forgalom érkezési folyamatát jellemző exponenseket aktív mérési módszerrel határozzuk meg. Megmutattuk, hogy a csomagvonalak diszperziója érzékeny a kiszolgálónál jelentkező torlódásra, valamint bevezettük a csomagvonal megnyúlását, mint a torlódott és nem torlódott állapot közötti fázisátalakulás rendparaméterét. Ezek használatával igazoltuk, hogy a megfelelő tulajdonságú próba forgalom indításával a két állapotot elválasztó kritikus pontban olyan adatok gyűjthetők, amelyekből meghatározhatóak a háttérforgalom érkezési folyamatát jellemző exponensek.
5. A bemutatott eredmények ellenőrzését és a bevezetett mérési módszerek vizsgálatát szimulációs módszerek mellett valódi rendszerekben is elvégeztük. Az aktív hálózati mérési módszerek, ezen belül a dolgozatban is hasz-

nált csomagpár és csomagvonat kísérletek valódi rendszerekben történő vizsgálatához kifejlesztettük, valamint megépítettük azokat a precíziós hálózati mérőgépeket amelyek együttese alkotja az Etomic Hálózati Mérési Infrastruktúrát. A mérési infrastruktúra mérőpontjait Európa szerte telepítettük.

Következtetések

- Az általunk kidolgozott granuláris modell lehetővé teszi a Internetes forgalom mérésekben használt diszperziós görbe leírását, ezáltal hálózati mérésekkel pontosan meghatározhatóvá teszi a hálózati útvonal főbb paramétereinek, azaz a fizikai sávszélességének, az aktuálisan ott lévő háttérforgalom átlagának, valamint a háttérforgalom granularitásának pontos értékét.
- Új módszert vezettünk be a csomagvonat diszperzió kiértékeléséhez. Ez lehetővé teszi, hogy egy adott hálózati útvonalon a háttérforgalom érkezési folyamatát aktív végpont-végpont mérésekkel határozzuk meg.

A tézisek alapjául szolgáló publikációk

- P. Hága, K. Diriczi, G. Vattay, I. Csabai:
Granular model of packet pair separation in Poissonian traffic
Computer Networks, Volume 51, Issue 3, Pages 683-698 (2007)
- P. Hága, S. Laki, F. Tóth, I. Csabai, J. Stéger, G. Vattay:
Neural Network Based Available Bandwidth Estimation in the ETOMIC Infrastructure
IEEE TridentCom 2007 Conference, **Candidate to Best Paper Award**,
21-23 May 2007, Orlando, FL, USA (2007)
- P. Hága, G. Vattay:
Detecting the Long-Range Dependence in the Internet Traffic with Packet Trains
European Conference on Complex Systems (ECCS'06), p181, 25-29 September 2006, Oxford, United Kingdom (2006)
- G. Simon, J. Stéger, P. Hága, I. Csabai, G. Vattay:
Measuring the Dynamical State of the Internet: Large Scale Network Tomography via the ETOMIC Infrastructure
Complexus 2006; Vol. 3, No. 1-3., Pages 119-130 (2006)

- P. Haga, K. Diriczi, G. Vattay, I. Csabai:
Understanding packet pair separation beyond the fluid model: The key role of traffic granularity
 IEEE INFOCOM 2006 Conference, Vol.III., Sec.50., 23-29 April 2006, Barcelona, Spain (2006)
- P. Haga, A. Pasztor, D. Veitch, I. Csabai:
PathSensor: Towards Efficient Available Bandwidth Measurement
 IPS-MoMe 2005 , p39-47, 14-15 March 2005, Warsaw, Poland (2005)
- D. Morato, E. Magana, M. Izal, J. Aracil, F. Naranjo, F. Astiz, U. Alonso, I. Csabai, P. Haga, G. Simon, J. Steger, G. Vattay:
ETOMIC: A testbed for universal active and passive measurements
 IEEE TridentCom 2005 Conference, **Best Testbed Award**, p283-289, 23-25 February 2005, Trento, Italy (2005)
- P. Haga, P. Pollner, G. Simon, I. Csabai, G. Vattay:
Self-generated Self-similar Traffic
 Journal of Nonlinear Phenomena in Complex Systems Vol.6, No.4, p814-823, (2003)