

MonCaptcha - Каптча Үйлчилгээ

П. Мөнхтулга*, Ч. Алтангэрэл*

*Монгол Улсын Их Сургууль. Хэрэглээний Шинжлэх Ухаан, Инженерчлэлийн Сургууль.

Мэдээлэл, Компьютерийн Ухааны Тэнхим
munkh87com@gmail.com

Abstract— Интернет нь хайлтын систем, электрон шуудан, нийгмийн сүлжээ зэрэг олон веб үйлчилгээнүүдийн тусламжтайгаар бидний амьдрал улам гүн нэвтэрч байна. Тэгэхдээ компьютерын програм болох бот нь интернет хэрэглэгчдэд байнга аюул учруулсаар ирсэн. Уг асуудлыг шийдвэрлэхийн тулд хэрэглэгчийг компьютерын программаас ялгах Каптчаг (хүн ба компьютеруудыг ялгахад зориулсан бүрэн автомат нээлттэй Тьюрингийн тест) хөгжүүлсэн. Хэдийгээр энэ механизм веб үйлчилгээнд автоматаар бүртгүүлэхээс хязгаарлах, сайн хамгаалалт боловч, англи хэл дээр байдаг нь Монгол хэрэглэгчдийн хувьд тохиромж муутай, мөн энэ үйлчилгээг үзүүлдэг сервер нь Монголоос гадна байрладаг нь зөвхөн Монгол улс дотроо локал байдлаар ашиглах боломжгүй. Ялангуяа хакеруудад Каптчаны механизмыг залилах боломжтой сул талууд зарим Каптчануудад байсаар байна. Тиймээс уг ажлын хүрээнд Бот-халдлагаас хамгаалах системийг хөгжүүлэхээр зорилоо.

Keywords— *Бот - Веб робот програм; CAPTCHA - КАПТЧА; workflow - үйл ажиллагааны диаграм; ERD - entity-relationship diagram; MySQL server; PHP; Apache server; MVC*

I. УДИРТГАЛ

Интернет нь хүний амьдралд харилцаа холбоо, боловсрол, онлайн худалдааны үйл ажиллагаа гэх мэт олон өөрчлөлтүүдийг авчирсан. Зарим веб үйлчилгээ болох Яахоо (Yahoo), Жимэйл (Gmail) ба Хотмэйл (Hotmail) - ууд нь хэрэглэгчдэд уг үйлчилгээг авахын тулд онлайнар бүртгүүлэх хэрэгтэй болдог. Хэдий тийм боловч хакерууд бүртгэлийн хуудсыг автоматаар бүртгэдэг програмыг бүтээсэн бөгөөд энэ нь бүртгэлийн хуудсанд хуурамч мэдээллийг их олон удаа оруулснаар системд ачаалал учруулж, түүний үйл ажиллагааг доод төвшинд хүртэл унагаж чадах ба үүнээс ч илүүгээр системд маш их бүртгэл орж ирснээр, түүнийг гацааж, зогсоох тохиолдолд хүргэдэг [2].

Тиймээс судлаачид онлайн бүртгэлийн үед хүн эсвэл компьютерын програм болохыг ялгах механизмыг хөгжүүлсэн. Өнөөдөр энэ асуудлыг шийдвэрлэхэд стандарт болсон механизм - Каптчаг (хүн ба компьютеруудыг ялгахад зориулсан бүрэн автомат нээлттэй Тьюрингийн тест) ашиглаж байна. Каптчаны гол санаа нь хүний онцлогт тулгуурлан боловсруулсан даалгаврыг хүн гүйцэтгэх бөгөөд компьютерын програм гүйцэтгэж чадахгүй байхаар хийгдсэн. Жишээ нь хэрэглэгчид харагдацыг гажуудуулсан бичвэрэн зураг эсвэл тодорхой тооны зургуудыг олон зургаас сонгох даалгавруудыг танилцуулдаг [2].

Сүүлийн үед олон төрлийн Каптча хөгжүүлсэн байна. Тэдгээрийн зарим нэг нь зурган тэмдэгт танигч (OCR) дээр суурилсан - Каптча [3], нөгөө хэсэг нь зурган тэмдэгт танигч биш (Non-OCR) мультимедиа буюу дуу-хоолойг ба видеог ашигласан.

Эдгээр Каптчаны төрлүүдээс зарим нэг нь шинэ бот-програмуудаар халдлагад өртөж байна. Жишээ нь Каптча текстийг үсгийн хэвийн сегментацын механизмаар тайлж болох юм. Энэ өгүүлэлд MonCaptcha буюу шинэ төрлийн Каптча үйлчилгээг танилцуулж байгаа бөгөөд өмнө хэвлүүлсэн өгүүлэлд монгол веб сайтуудын каптча хэрэглээний дүн, каптча системийн хөгжүүлэлтийн тухай өгүүлсэн.

II. ХОЛБООТОЙ АЖЛУУД

A. Каптчаны тодорхойлолт

Каптча бол веб сайтыг ботуудаас хамгаалдаг бөгөөд компьютерын програм хариулж чадахааргүй, харин ганц хүн хариултыг хийж чадах тестийг хийдэг ба үнэлдэг програм юм [1]. Жишээ нь зураг 1 - д байгаа гажуудуулсан бичвэрийг хүн уншиж чадах ба компьютер түүнийг уншиж чадахгүй.



Figure 1. Каптчаны хэлбэрүүд.

B. Каптча хэрэглээ

- 1) Спам мэдээллээс хамгаалах: Блог болон веб сайтыг коментын (веб сайт хэрэглэгчид сэтгэгдэл, саналаа үлдээдэг хэсэг) хэсэгт хэрэгцээгүй зар сурталчилгааны зорилготой мэдээлэл оруулахаас хамгаалах [2,5].
- 2) Веб сайтыг бүртгэлийг хамгаалах: Каптча зарим нэг үнэ төлбөргүй Hotmail, Gmail, Yahoo зэрэг веб үйлчилгээг минутанд хэдэн мянган хэрэглэгч бүртгэж чадах автомат бүртгэлийн ботоос хамгаалдаг [2,5].
- 3) Scraper - уудаас имэйл хаягийг хамгаалах: Интернетэд электрон шууданг хайдаг спам төрлийн бот нь электрон шууданг интернет веб сайтуудаас автоматаар цуглуулж спам мэдээлэл илгээдэг бөгөөд үүнээс сэргийлэхийн тулд веб сайтууд электрон шуудангаа харуулах даа каптча бөглөх байдлаар шийдэж болдог [2,5].

- 4) Санал асуулгыг хамгаалах: Зарим веб сайтуудад хэрэглэгчийн саналыг судлах, цуглуулах, санал хураалт явуулах зэрэг хэрэгцээ шаардлага тулгардаг энэ үед автомат програмын тусламжтайгаар буруу санал хураалтад нөлөөлөх боломжтой байдаг бөгөөд үүнээс хамгаалах нэг арга нь каптча юм. Каптча нь хэрэглэгчийг санал өгүүлэхийн тулд заавал бүртгэлтэй байхыг шаардахгүй [2,5].
- 5) Dictionary халдлагаас хамгаалах: Веб сайтын нэвтрэх хуудсыг автомат паспорт оруулах аргаар довтолдог бөгөөд үүнээс Каптча хамгаалж өгдөг. Энэ төрлийн халдлагыг Dictionary Attack /Толь-бичгэн халдлага/ гэдэг [2,5].
- 6) Хайлтын ботоос хамгаалах: Заримдаа веб сайтуудын хувьд веб сайтын зарим хуудаснуудыг unindexed буюу хэрэглэгч тухайн хуудсыг хялбар олохоос сэргийлэх шаардлага байдаг. Тиймээс html - ийн зориулалтын тагийг хэрэглэж болох ба энэ нь хайлтын ботоос хамгаална гэсэн баталгаа өгч чаддаггүй. Үүнээс сэргийлэх илүү үр дүнтэй арга бол Каптча юм. Энэ төрлийн ботын халдлагыг "Search Engine Bots" аргаар хийдэг [2,5].

С. Зааварчилгаа

Үйлчлүүлэгч веб сайт ботоос өөрийн сайтыг хамгаалахын тулд каптчаг хэрэглэдэг. Хэрэглэгчийн шаардлагаас хамаарч каптча нь олон төрөл зүйл байдаг бөгөөд дараах зааварчилгааг бүх төрлийн каптчад зөвлөмж болгож байна [4]. Үүнд:

- Каптча хэрэглэхэд хялбар байх ёстой. Каптча нь зөвхөн бичвэр суурьт, зурган суурьт байж ботоос хамгаалахаас гадна харааны бэрхшээлтэй иргэд хэрэглэхэд хүндрэлтэй. Иймээс ямар нэгэн каптча хараагүй иргэдэд үйлчлэх дуу, хоолойн бичлэгийг танилцуулах хувилбартай байх ёстой [4].
- Зурган суурьт каптча нь автомат халдлагаас хамгаалах үүднээс зурган бичвэрийг гажуудуулсан (бичвэрийн хэлбэр дүрс эвдэрсэн) байх шаардлагатай.
- Каптчаг байгуулагч код нь хамгаалагдсан хялбар биш байх ёстой. Тиймээс дүрслэгдсэн зураг компьютероор унших боломжгүй байх ёстой ба систем скриптийн төвшинд уг тестийг залилах боломжгүй байдлыг хангасан байна. Энгийн жишээ дурдвал, (1) системүүд Каптчаны хариуг энгийн веб-формын нэг хэсэг мэт дамжуулах. (2) систем нэг төрлийн Каптчаг олон дахин асуух (энэ нь Каптчаг "replay халдлага"-д өртөмхий болгоно). Интернетэд нээлттэй эхтэй олдох ихэнх Каптча скриптүүд нь түүнийг өөрийг нь

энэ төрлийн халдлагад өртөмхий болгож байдаг [4].

- Каптча нь их хэмжээний сайтууд хэрэглэх тохиолдолд аюулгүй ажиллагаатай байх ёстой.

reCaptcha систем нь 2000 онд бүтээгдсэнээсээ хойш өнөөдрийг хүртэл өргөн хэрэглэгдэж байгаа бөгөөд дээрх бүх шаардлагыг хангадаг [4].

D. reCaptcha сервис үйлчилгээ

Зарим гар бичвэрийг програмаар тайлан таних боломжгүй байдаг. Тиймээс капчанд нэмж асуулга байдлаар тэр бичвэрийн зургийг оруулснаар шийддэг. Үүнийг reCaptcha анх хэрэгжүүлж эхэлсэн [3].

reCAPTCHA үйлчилгээг веб сайтдаа ашиглахдаа дараах үйлдлийг дэс дараалан гүйцэтгэнэ.

- 1) Хэрэглэгч сервис үйлчилгээг авахын тулд Google - д бүртгүүлнэ, хэрвээ бүртгүүлсэн бол хэрэглэгчийн эрхээрээ reCAPTCHA авах гэсэн дээр дарж орно. [<https://www.google.com/recaptcha/admin>]
- 2) Шинэ сайт бүртгүүлэх хэсэгт тайлбар, домэйн нэр, электрон шуудангийн хаягийг оруулан бүртгүүлнэ. Үүний дараа site key ба secret key-г үүсгэн харуулна. Нээлттэй түлхүүр (site key) - нь хэрэглэгчид зориулсан, ил түлхүүр юм. Хувийн түлхүүр (secret key) - нь хэрэглэгч сайт болон reCaptcha сервисийн хооронд холбоо тогтооход хэрэглэх бөгөөд аюулгүй байдлын үүднээс нууцлах ёстой түлхүүр юм [6].
- 3) Хэрэглэгч веб сайтын клиент ба серверийн талбаруудад reCaptcha сервис үйлчилгээний зааврын дагуу тохиргоо хийнэ [6].
- 4) Клиент талын тохиргоонд Каптчаг дүрслэх скрипт код ба тагууд хийгдэнэ [6].
- 5) Веб сайтын каптчаг суурилуулсан формыг ажиллуулахад reCAPTCHA сервертэй холболт хийгдэж хүсэлт илгээн хариуд нь Каптча дэлгэц дээр дүрслэгдэнэ [6].
- 6) Сервер талын тохиргоонд веб сайт, reCAPTCHA хоёрын хоорондын авах, өгөх урсгалын өгөгдлийн боловсруулалтын тохиргоо хийгдэнэ [6].

III. СИСТЕМИЙН ХЭРЭГЖҮҮЛЭЛТ БА ТЕХНОЛОГИ

E. MonCaptcha системийн үйл ажиллагааны диаграм

Зураг 2 - т MonCaptcha системийн объектын үйл ажиллагааг ерөнхийд нь дараах диаграммаар байгуулсан бөгөөд reCaptcha системийн үйл ажиллагаанд суурилсан.

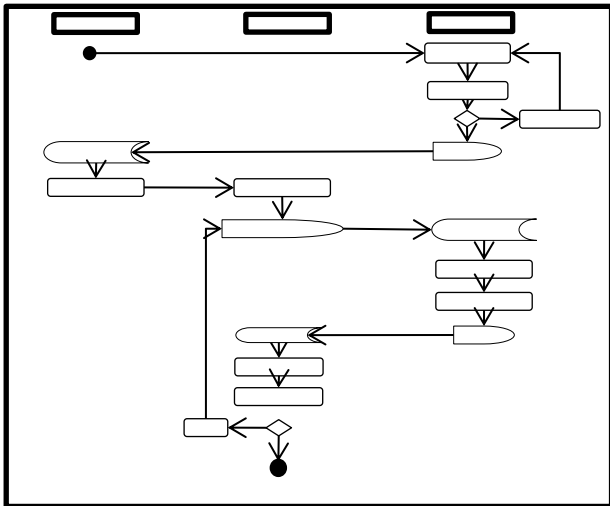


Figure 2. Үйл ажиллагааны диаграм.

Каптча үйлчилгээ авах хэрэглэгч өөрийн веб сайтын домэйн нэрийг бүртгүүлж, имэйл хаягаар нь ирэх зааварчилгааны дагуу өөрийн нууц үгийг авч, веб сайтдаа тохиргоо хийснээр ашиглах боломжтой болох юм.

F. Өгөгдлийн сангийн хэрэгжүүлэлт

Зураг 3 - т Каптча үйлчилгээний ерөнхий ER диаграммыг дүрслэв. Үүнд userst хүснэгт хэрэглэгч веб сайтын мэдээллийг агуулна, smart_answer хүснэгт ухаалаг хариултыг агуулна, зурган бичвэрийн мэдээлэл гэх мэтээр агуулна.

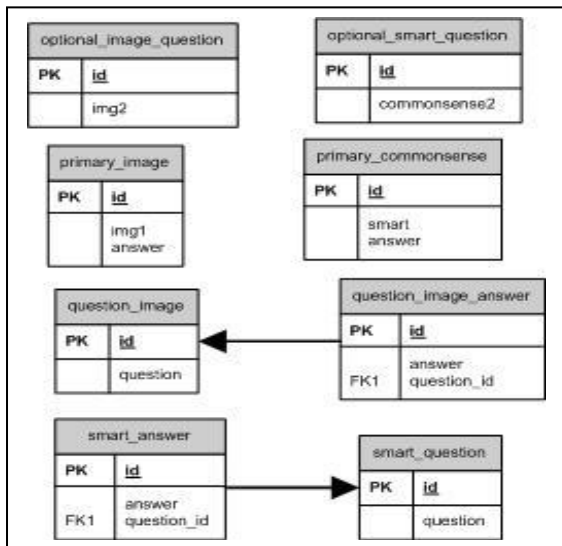


Figure 3. Өгөгдлийн сангийн холбоост загвар

Зураг 4 - т MonCaptcha системийн хэрэглэгч бүртгүүлэх интерфэйсийг дүрслэв. Үүнд хэрэглэгч (1) бүртгүүлэх веб сайтын домэйн нэр, (2) Холбоо барих утасны дугаар, (3) Хэрэглэдэг имэйл хаягаар бүртгүүлнэ.

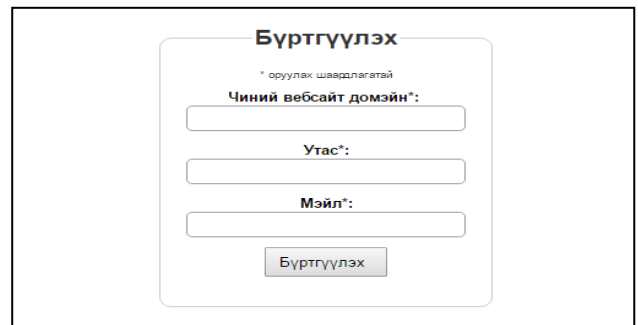


Figure 4. Хэрэглэгч бүртгүүлэх форм.

Каптча нь хэрэглэгчээс ухаалаг асуулт болон зурган гар бичвэрийн асуулга асуух ба хэрэглэгч асуулгыг хариулж зөвшөөрөх товчлуурыг дарна. MonCaptcha үйлчилгээ ухаалаг асуултын хариуг дүгнэх ба буруу бол дүгнэхгүй. Энэ асуултын гол санаа нь хүний ойлголтод дээр суурилсан. Систем нь хүний ойлголттой холбоотой асуулт асуух бөгөөд зөвхөн хүн хариулах боломжтой.

Зураг 5 - т MonCaptcha системийн Каптча асуултын хэрэглэгчийн интерфэйсийг дүрслэв.

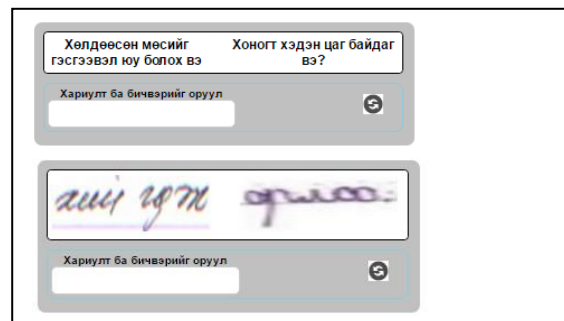


Figure 5. MonCaptcha системийн хэрэглэгчийн интерфэйс

G. Технологи

PHP бол веб хөгжүүлэлтэд зориулагдан хийгдсэн, сервер талын програмчлалын хэл юм [7, 9]. Facebook, Twitter болон бусад сайтууд уг хэлийг сонгон хэрэглэдэг [7, 9] тул PHP хэлийг сонгов.

MySQL нь нээлттэй эх кодтой, өгөгдлийн сангийн менежмент систем бөгөөд кроссплатформ програм хангамж юм. Мөн PHP хэл MySQL хэлтэй хослон ажиллах боломжтой [16] ба веб програмчлалд өргөн хэрэглэдэг учир MySQL - г сонгов.

Apache веб сервер нь UNIX ба Windows зэрэг олон үйлдлийн системд суурилан ажилладаг бөгөөд 1995 оноос анх олон нийтэд хэрэглэгдэж эхэлсэн, хамгийн алдартай веб сервер юм [10].

Серверийг Ubuntu 64 bit LTC 16.04 үйлдлийн системээр байгуулсан. Ubuntu бол линүкс үйлдлийн систем бөгөөд нээлттэй эхийн програм хангамж юм [11].

Модель-Харагдац-Удиртгал (MVC) нь програм хангамжийн архитектур бөгөөд програм хангамжийн инженерчлэлд ашиглагддаг загвар юм [7]. Иймээс MonCaptcha системийн хөгжүүлэлтийг MVC архитектураар зохион байгуулсан.

ҮР ДҮН

Системийг онлайн сервер дээр байршуулж хэрэгжүүлсэн. Ухаалаг асуулгын хэмжээ нь 10 ширхэг, харин зурган гар бичвэрийн асуулгын хэмжээ нь 398 ширхэг байна.

MonCaptcha системийг Вебсайтын програмчлалыг PHP объект хандалтат програмчлалын хэлээр бүтээж, үндсэн загварыг HTML, CSS, Javascript гэсэн 3 хэлний тусламжтайгаар хийсэн. Ингэж хийхдээ хэрэглэгчийн талд каптча форм зурах зориулалтаар шинээр Javascript санг бүтээсэн. Энэ сан нь хэрэглэгчийн талд дуудагдан ажиллах зориулалттай MonCaptcha системийг зурах скрипт юм.

Үр дүнд HTML, CSS, JavaScript, PHP хэл дээр 30 класс, 4164 мөр код бүхий, Crypt/RSA (PHPSECLIB сангийн RSA сан), jQuery.js санг ашигласан MVC архитектуртай систем хөгжүүлэв.

Дараах зурагт MonCaptcha системийн нэвтрэх хэсэгт каптча харагдах байдал.

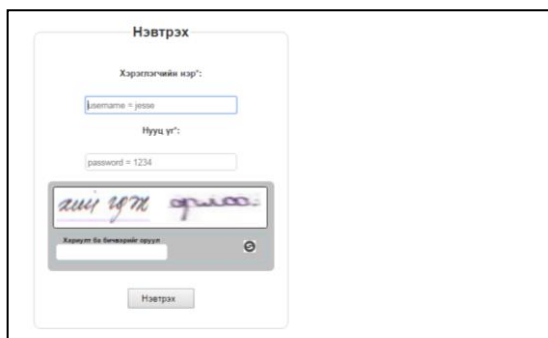


Figure 6. Каптча үйлчилгээг хэрэглэсэн нэвтрэх хуудас

ДҮГНЭЛТ

Каптча нь веб сайтыг ботоос хамгаалах гол ач холбогдолтой систем юм. Энэ өгүүлэлд системийн хэрэгжүүлэлттэй холбоотой ажлуудыг тодорхойлсон ба системийн хэрэгжүүлэлт ба технологийг өгүүлсэн. MonCaptcha системийн хувьд онлайн сервер дээр байршуулсан, дараах холбоосоор <http://172.104.34.197/MonCaptcha/> үзэх боломжтой.

НОМ ЗҮЙ

- [1] Carnegie Mellon University, CAPTCHA: Telling Humans and Computers Apart Automatically. Available from: <http://www.captcha.net/> [Accessed: March 28, 2017].
- [2] Walid Khalifa Abdullah Hasan "A SURVEY OF CURRENT RESEARCH ON CAPTCHA" in International Journal of Computer Science & Engineering Survey (IJCSES) Vol.7, No.3, June 2016
- [3] Luis von Ahn, Benjamin Maurer, Colin McMillen, "reCAPTCHA: Human-Based Character Recognition via Web Security Measures", in Computer Science Department, Vol. 321., 2008.
- [4] Carnegie Mellon University, CAPTCHA: Guidelines. Available from: <http://www.captcha.net/> [Accessed: March 12, 2017]

- [5] Carnegie Mellon University, CAPTCHA: Applications of CAPTCHAs. Available from: <http://www.captcha.net/> [Accessed: March 12, 2017]
- [6] Apply Google reCAPTCHA into Your Site. Available from: <https://www.codeproject.com/Articles/1063051/Apply-Google-reCAPTCHA-into-Your-Site> [Accessed: March 11, 2017]
- [7] Bartosz Porebcki, Karol Przystalski, Leszek Nowak., "Building PHP Applications", 2011
- [8] Wikipedia The Free Encyclopedia, Available from: <https://en.wikipedia.org/wiki/MySQL> [Accessed: March 11, 2017]
- [9] Wikipedia The Free Encyclopedia, Область применения. Available from: <https://ru.wikipedia.org/wiki/PHP> [Accessed: March 11, 2017]
- [10] Apache http server project, <http://httpd.apache.org/>.
- [11] Ubuntu (operating system). [https://en.wikipedia.org/wiki/Ubuntu_\(operating_system\)](https://en.wikipedia.org/wiki/Ubuntu_(operating_system)) [Accessed: March 11, 2017]