(Borang JKPTA FTSM UKM 3)

UNIVERSITI KEBANGSAAN MALAYSIA
National University of Malaysia

FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT

BORANG PENYERAHAN LAPORAN ILMIAH

SEM 2 **SESI** 2021 / 2022

Bahagian A: Maklumat Diri Pelajar
Part A: Student's Details

No. Matrik *(Matric Number)*	A182536
Nama *(Name)*	Dinie Syahirah binti Zairol Sahrin
Program pengajian (*Programme*)	Teknologi Maklumat
No. Telefon (*Telephone Number*)	01119999718
Emel (*Email*)	a182536@siswa.ukm.edu.my

Tajuk Projek (*Project Title*):

Sistem Sambungan Web bagi Pencegahan Pancingan Data (SeaPol)

Tandatangan (*Signature*): _____ Tarikh (*Date*) : 22 Julai 2022

Bahagian B: Perakuan Penyelia
Part B: Supervisor's Approval

Saya peraku laporan ini telah disemak dan dibaiki, dan **menyokong / tidak menyokong*** penyerahan laporan ilmiah ini.
*I certify that this report has been reviewed and amended, and **approved / rejected*** the report submission.*

Tandatangan (*Signature*): _____ Tarikh (*Date*) : 25 Julai 2022

Cap Rasmi :
(Official Stamp)

MASNIZAH MOHD (PhD)
Associate Professor
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia

SISTEM SAMBUNGAN WEB BAGI PENCEGAHAN PANCINGAN DATA (SEAPOL)

DINIE SYAHIRAH ZAIROL SAHRIN
MASNIZAH MOHD

Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia

ABSTRAK

Pancingan data ialah kaedah mengumpul maklumat peribadi yang berharga menggunakan e-mel atau laman web yang mengelirukan pengguna. Matlamatnya adalah untuk menipu pengguna supaya mempercayai bahawa mesej dapati itu adalah sesuatu yang mereka mahu atau perlukan. Masalah ini dapat dinyatakan kerana pancingan data masih merupakan salah satu isu yang paling meluas dan berniat jahat, dengan mesej dan teknik pancingan data menjadi semakin canggih. Dari hasil masalah yang diselidiki oleh pengguna yang lain, didapati bahawa kebanyakkan orang dari semua latar belakang dan usia masih sukar untuk membezakan antara laman web yang sebenar dan yang palsu sehingga menjadi mangsa aktiviti pancingan data. Oleh itu, objektif kajian diwujudkan bagi memahami pertumbuhan dalam aktiviti pancingan data ini yang mana boleh meningkat dengan pesat mengikut peredaran masa. Objektif tersebut boleh diukur berdasarkan mengenal pasti unsur pancingan data dalam talian, mereka bentuk sambungan yang mengesan aktiviti pancingan data dan juga menilai keberkesanan sistem sambungan tersebut. Untuk membangunkan projek ini, metodologi yang akan digunakan adalah kaedah *agile* dimana kaedah ini bersifat fleksibel dan mudah untuk digunakan kerana ia tidak tertumpu kepada satu fasa sahaja dan tidak tetap. Berdasarkan hasil penyelidikan yang didapati, sistem sambungan web bagi pencegahan pancingan data yang diberi nama SeaPol telah boleh digunapakai dan berfungsi sebaiknya untuk mengesan sebarang aktiviti pancingan data dalam talian. Hasil ujian sistem ini juga direkod dengan para penguji diberi peluang untuk menguji tahap kebolehgunaan sistem tersebut. Dengan terlahirnya sistem sambungan web bagi pencegahan pancingan data (SeaPol) ini maka ia dapat memberi sumbagan kepada para pengguna di mana mereka dapat membezakan laman web yang mempunyai unsur pancingan data ataupun sebaliknya. Melalui cara ini, para pengguna dapat menyelamatkan diri mereka daripada menjadi mangsa pancingan data dan sekaligus dapat menyebarkan kesedaran kepada orang ramai mengenai bahaya aktiviti pancingan data tersebut. Kesimpulan yang diharapkan adalah orang ramai dapat memahami dan mengambil tahu dengan lebih lanjut tentang

bagaimana para pengguna internet dari semua peringkat umur dan latar belakang boleh menjadi mangsa taktik jika tidak digunakan dengan betul.

1 PENGENALAN

Internet telah berkembang di seluruh dunia dan diterima pakai secara meluas sehingga dunia dikenali sebagai dunia tanpa sempadan. Dengan perkembangan pesat rangkaian komputer tanpa wayar, seperti Wi-Fi, pengguna semakin terdedah kepada serangan dan ancaman terhadap sistem komputer mereka. Masalah bermula apabila semakin meningkat bilangan kes yang melibatkan salah laku penggunaan komputer seperti pancingan data dalam emel atau laman web. Ini boleh ditunjukkan oleh kekurangan pengetahuan mengenai pancingan data kepada orang ramai kerana mereka adalah orang yang mungkin tidak tahu bahawa data mereka diambil tanpa disedari untuk tujuan jahat. Selain itu, masalah ini boleh dikenal pasti kerana kurang kesedaran orang ramai tentang pancingan data dan kesannya terhadap kehidupan orang ramai.

Satu praktis untuk mengelirukan pengguna-pengguna Internet (melalui penggunaan mesej emel atau laman web menipu) ke dalam mendedahkan maklumat peribadi atau sulit yang mana ia boleh digunakan secara haram. Definisi pancingan data secara mendalam, Phish ini disebut sama seperti ia dieja, iaitu mengatakan seperti perkataan "ikan". Konsep ini boleh digambarkan sebagai seorang nelayan yang melempar mata kail berumpan di luar sana dan berharap orang ramai akan menggigitnya. Pancingan data ialah kaedah cuba mengumpul maklumat peribadi yang berharga menggunakan e-mel atau tapak web yang mengelirukan dengan niatnya adalah untuk menipu dan mengelirukan pengguna supaya mempercayai bahawa mesej itu adalah sesuatu yang mereka mahu atau perlukan. Pancingan data merupakan salah satu taktik yang paling meluas dan dijalankan sehingga kini, dengan mesej dan teknik pancingan data menjadi semakin canggih saban hari. Dalam bab ini, akan dibentangkan penjelasan ringkas tentang projek yang akan dijalankan. Menerangkan aspek utama seperti pernyataan masalah, objektif masalah dan skop projek.

2 PENYATAAN MASALAH

Pancingan data kadangkala boleh menyebabkan kerosakan beribu malah berjuta ringgit setiap tahun dan memberi ancaman serius kepada pengguna Internet. Walaupun telah banyak kesedaran pancingan data diadakan, kebanyakan pengguna masih tidak dapat menentukan dan tidak tahu sama ada laman web yang mereka gunakan mempunyai niat jahat atau tidak sehinggan boleh menyebabkan mereka terjerumus ke dalam perangkap penyerang. Ini kerana unsur pancingan data dalam perisian atau alatan sedia ada masih belum dikemaskini seperti senarai laman web kategori malicious dan suspicious.

3 OBJEKTIF KAJIAN

Objektif utama yang boleh dikaitkan dengan penyataan masalah yang telah dilampirkan ialah:

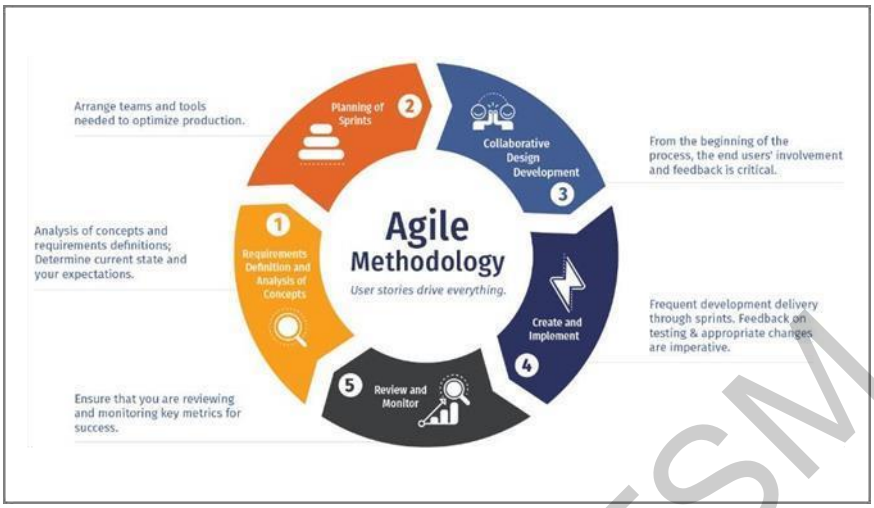
i. Untuk mengenal pasti unsur pancingan data dalam talian.

ii. Untuk mereka bentuk sebuah sambungan atau *plug in* yang mengesan aktiviti pancingan data dalam talian.

iii. Untuk menilai keberkesanan pendekatan yang direka bentuk dalam mengesan pancingan data.

4 METOD KAJIAN

Dalam pembangunan projek ini, model proses yang akan digunakan ialah kaedah Agile. Agile merupakan istilah yang dicipta oleh AgileAlliance, sekumpulan profesional perisian Agile (Mark.C Paulk, 2002) Matlamat utama konsep membangunkan perisian Agile ini adalah untuk memuaskan pengguna atau pelanggan melalui penghantaran awal projek, iaitu tempoh awal yang telah ditetapkan dan pengguna atau pelanggan boleh mengubah permintaan mereka bila-bila walaupun perisian sudah berada dalam fasa akhir. (Mark.C Paulk,2002). Selain itu, pengguna atau pelanggan dan pembangun haruslah bekerjasama sepanjang projek dijalankan untuk memastikan benda yang di bangun menepati permintaan pengguna atau pelanggan. Proses Agile juga menggalakkan pembangunan mampan. Berikut merupakan rajah 1.1 bagi metodologi Agile.



Rajah 4.1: Model Agile (Sumber: https://www.nvisia.com)

i. Fasa Analisis *(Analysis)*

Fasa ini bertujuan untuk menganalisis keperluan dan perisian bagi membangunkan sambungan dalam *chrome* 'SeaPol'. Dalam fasa ini pengumpulan data dijalankan untuk memenuhi kehendak keperluan dan kehendak pengguna.

ii. Fasa Perancangan *(Planning)*

Fasa ini dilaksanakan dengan memfokuskan kepada perancangan dan persiapan kitaran projek. Segala bentuk persiapan seperti mengatur kumpulan dan alat-alat perisian yang akan digunapakai diadakan untuk mengoptimumkan dan menambahbaik produksi supaya hasil yang terbaik dapat dikeluarkan dengan baik.

iii. Fasa Pembangunan *(Development)*

Pembangunan sistem dilaksanakan dalam fasa ini. Reka bentuk dihasilkan bagi sambungan *chrome* SeaPol mestilah sesuai dan menepati keperluan serta mudah difahami oleh pengguna. Dengan mereka bentuk sistem sambungan web dimana ia akan mula mengesan laman web pancingan data dengan menetapkan peraturan dan aturan tertentu, hasilnya apabila laman web pancingan data dikesan, sistem akan mengeluarkan pemberitahuan *pop-up* mengatakan laman web tersebut berunsur pancingan data. Peraturan yang diterapkan dalam sistem adalah berdasarkan capaian URL dan alamat IP dimana satu set aturan yang ditetapkan, contohnya `if(((url.substring(url.search(srch1))).match(srch2))==null)` maka ia akan di identifikasikan sebagai NP *(non-phishing)* kerana *variable* yang ditemui adalah padan dengan aturan tersebut dan jika sebaliknya ia akan diklasifikasikan sebagai P *(phishing).*

Variable srch1 adalah menentukan simbol "//" dan srch2 adalah menentukan "https". Setiap capaian URL dan alamat IP ditapis konfidentalitinya mengikut set aturan yang ditetapkan dan pembangunan sistem haruslah selari dengan peraturan-peraturan yang telah ditetapkan.

iv. Fasa Pelaksanaan *(Implement)*

Fasa ini merupakan fasa di mana projek dibina menggunakan bahasa pengaturcaraan dan pangkalan data yang dipilih. Dalam fasa ini, pelaksanaan fungsi sambungan *chrome* 'SeaPol' dibina mengikut objektif.

v. Fasa Pengujian dan Semakan *(Testing and Review)*

Fasa ini merupakan fasa yang penting dalam metodologi ini. Hal ini kerana ralat yang terdapat di dalam penyambung krom 'SeaPol' dapat dikenal pasti dan diperbaiki dengan segera disebabkan oleh pengujian yang berterusan pada setiap peringkat.

5 HASIL KAJIAN

Fasa pembangunan merupakan fasa di mana aktiviti pengekodan, antara muka serta fungsi lain-lain akan dilaksanakan. Pembangunan sistem seharusnya menepati ciri-ciri sistem yang telah ditetapkan dalam fasa reka bentuk supaya sistem yang dibangunkan dapat mematuhi keperluan pengguna serta mencapai objektif. Untuk mencapai objektif projek, beberapa fungsi yang diperlukan telah pun dibina dan dimasukkan ke dalam sistem bagi Sistem Sambungan Web bagi Pencegahan Pancingan Data ini. Antara fungsi yang telah diimplementasikan ke dalam sistem ini adalah mengesan laman web pancingan data serta menyekat laman web pancingan data tersebut.

Untuk mencapai objektif sistem ini iaitu untuk mengenal pasti unsur pancingan data dalam talian, maka sebuah sambungan atau pemasang yang mengesan aktiviti pancingan data dalam talian telah direka untuk menilai keberkesanan pendekatan yang direka bentuk dalam mengesan pancingan data. Semasa fasa pembangunan sistem, operasi pengekodan dijalankan untuk mencipta sistem yang memenuhi ciri dan objektif projek. Fasa ini perlu dijalankan selaras dengan spesifikasi yang memenuhi kehendak dan keperluan pengguna.

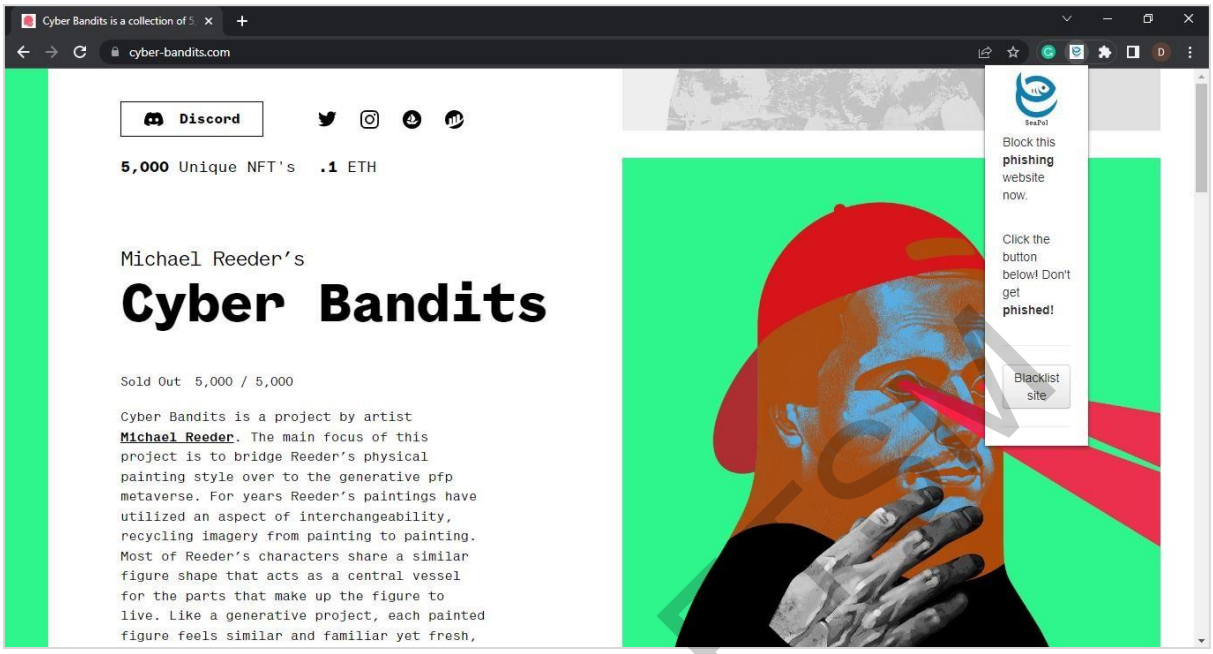
5.1 Mengesan laman web pancingan data

Rajah 5.1 menunjukkan antara muka bagi sistem sedang mengesan web yang berunsur pancingan data. Apabila sistem selesai mengesan web pancingan data tersebut, ia akan mengeluarkan pemberitahuan secara pop-up mengatakan bahawa laman web tersebut mempunyai unsur pancingan data.

Rajah 5.1: Antara muka ketika sistem sedang mengesan laman web.

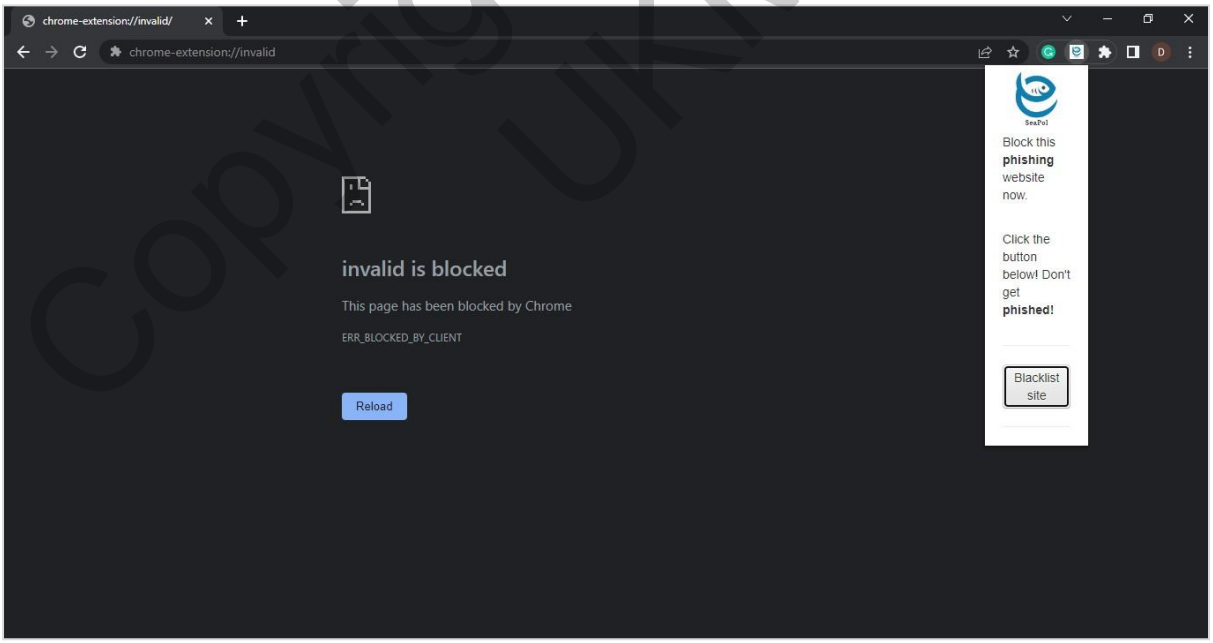
5.2 Menyekat laman web pancingan data

Rajah 5.2 menunjukkan laman web yang telah dikesan mempunyai unsur pancingan data. Butang 'Blacklist site' dipapar untuk pengguna menyekat laman web pancingan data tersebut.

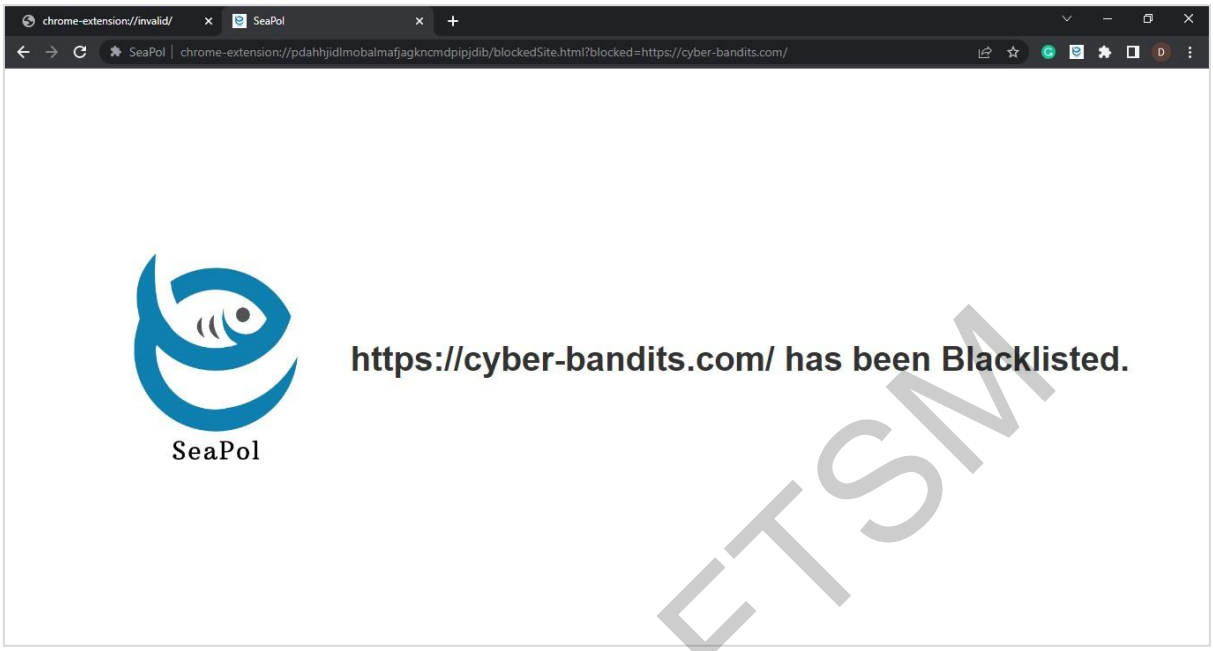


Rajah 5.2: Laman web yang telah dikesan.

Rajah 5.3 dan rajah 5.4 menunjukkan laman web pancingan data tersebut telah disekat oleh sistem. Laman web pancingan data tersebut tidak lagi boleh dibuka pada masa akan datang.

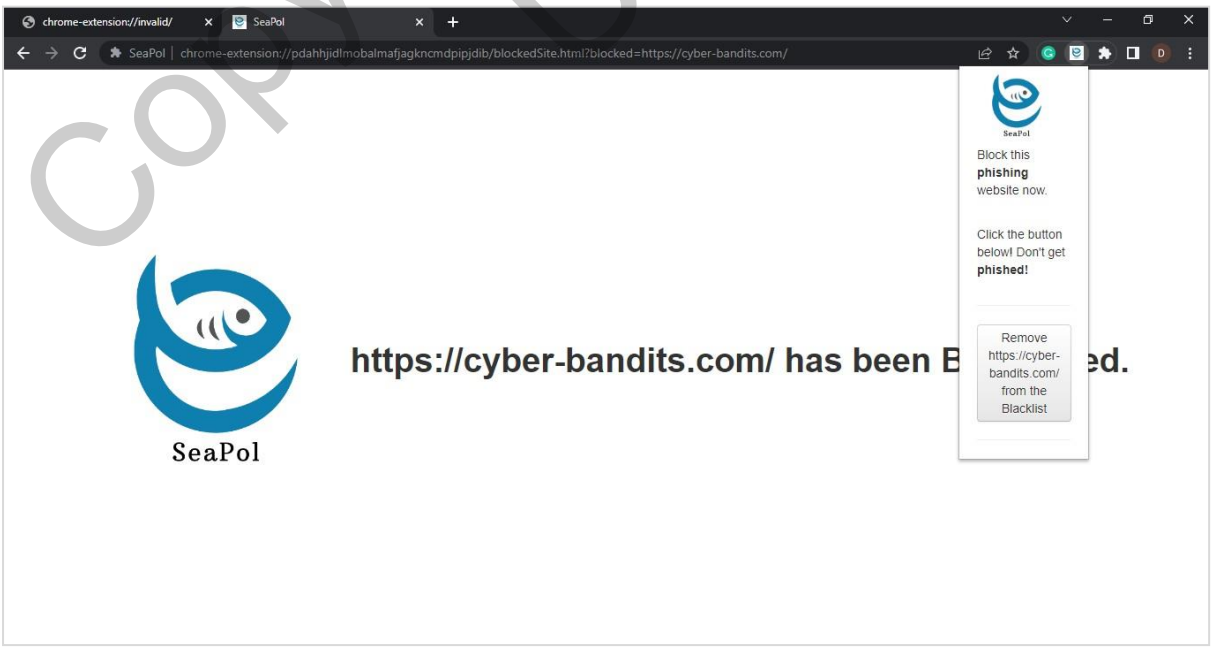


Rajah 5.3: Laman web setelah disekat oleh sistem.

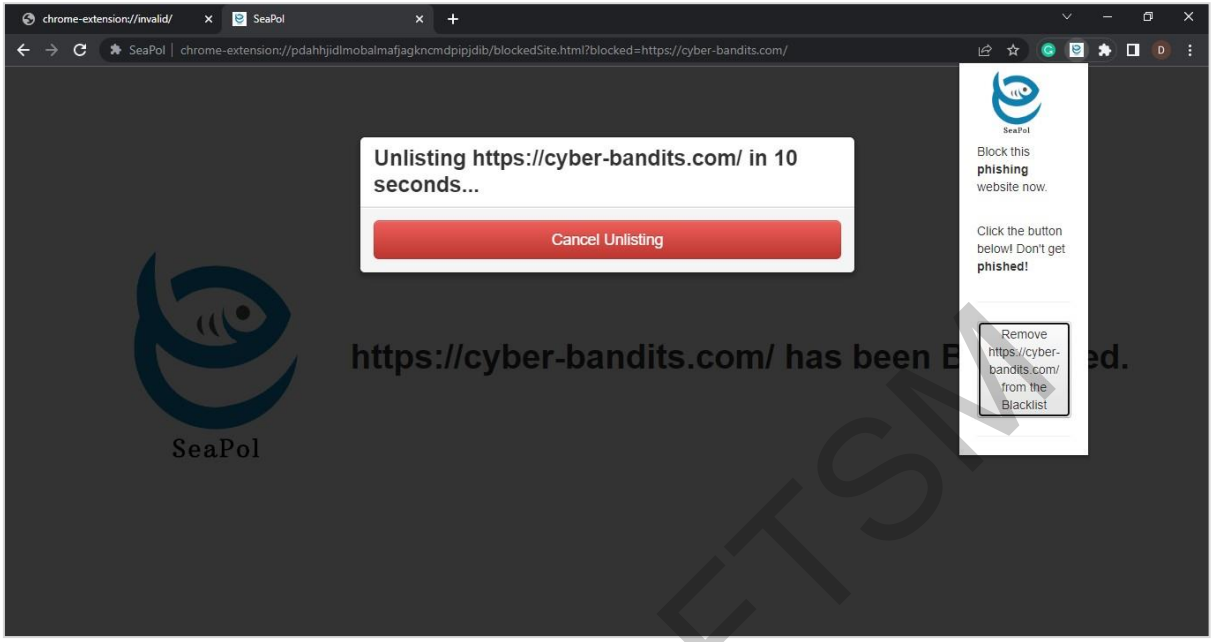


Rajah 5.4: Sistem menyatakan laman web telah disekat oleh sistem.

Rajah 5.5 dan rajah 5.6 menunjukkan ciri sekat laman web pancingan data boleh di patah balik keputusannya untuk tidak menyekatnya kembali. Pengguna akan menekan butang 'Remove website form the blacklist' untuk menukarkan status laman web tersebut daripada disenarai hitam.



Rajah 5.5: Menukar ciri sekat kepada tidak sekat.



Rajah 5.6: Sistem sedang menukar fungsi sekat kepada tidak sekat.

5.3 Hasil pengujian kebolehgunaan

Maklum balas yang dikumpulkan dari 3 pengguna melalui soal selidik kebolehgunaan telah dianalisis. Jadual 5.1 dan 5.2 menunjukkan demografi pengguna.

Jadual 5.1 Umur Responden

Umur	Kekerapan	Peratus (%)
20 tahun ke bawah	0	0
21 - 30 tahun	5	100
30 tahun ke atas	0	0

Jadual 5.2 Tahap Pendidikan responden

Umur	Kekerapan	Peratus (%)
Pendidikan Menengah	0	0
Pendidikan Pra-Universiti	3	60
Pengajian Tinggi	2	40

Merujuk kepada Jadual 5.1, semua responden berumur dari 21 hingga 30 tahun, iaitu 100% jumlah keseluruhannya. Responden yang berumur bawah 20 tahun serta 30 tahun ke atas mencatatkan 0% keseluruhannya. Jadual 5.2 pula menunjukkan majoriti daripada responden (60.0%) mempunyai tahap pendidikan Pra-Universiti manakala (40%) adalah daripada tahap

pendidikan pengajian tinggi. Setiap soalan soal selidik kebolehgunaan mempunyai 5 pilihan iaitu “Sangat Tidak Setuju”, “Tidak Setuju”, “Agak Setuju”, “Setuju”, dan “Sangat Setuju” dan markah mereka adalah dari 1 hingga 5. Markah yang diberi oleh setiap responden direkod dan digunakan untuk membuat analisis.

a) Kemudahan Kegunaan

Berdasarkan jadual 5.3, aspek pertama yang diuji ialah tahap kemudahan kegunaan sistem kepada pengguna. Soalan dalam bahagian ini bertanya tentang pendapat pengguna terhadap tahap kemudahan semasa menggunakan sistem. Nilai purata dalam bahagian ini adalah 1. Kumpulan data menunjukkan kebanyakan pengguna bersetuju bahawa sistem ini adalah mudah untuk digunakan dan ini boleh dilihat daripada nilai purata keseluruhan yang dijana dari nilai purata bagi setiap soalan.

Jadual 5.3 Kemudahan Kegunaan

Soalan	Median
Kemudahan Kegunaan	**4.5**
1. Saya dapat menggunakan sistem ini tanpa panduan.	**4.5**
2. Saya dapat memahami apa yang berlaku sepanjang menggunakan sistem.	**4.5**
3. Saya dapat menyekat laman web yang mempunyai unsur pancingan data dengan mudah.	**4.5**
4. Sistem ini dapat mengesan laman web yang berunsur pancingan data dengan cepat.	**4.5**
5. Sistem ini menunjukkan arahan yang jelas dan mudah difahami.	**4.5**

b) Kepuasan Antara Muka

Berdasarkan Jadual 5.4, faktor yang diuji seterusnya ialah untuk menguji tahap kepuasan antara muka keseluruhan sistem. Kebanyakan pengguna bersetuju bahawa mereka berpuas hati dengan antara muka aplikasi.

Jadual 5.4 Kepuasan Antara Muka

Soalan	Median
Kepuasan Antara Muka	**4.5**
1. Skema warna sistem yang digunakan adalah sesuai dan menarik.	**4.5**
2. Jenis huruf dan ukuran perkataan yang digunakan adalah sesuai.	**4.5**
3. Antara muka pengguna sistem ini konsisten.	**4.5**
4. Reka bentuk sistem ini adalah menarik.	**4.5**
5. Saya berpuas hati dengan antara muka sistem ini.	**4.5**

c) Kebolehgunaan

Berdasarkan Jadual 5.5, menguji tahap kebolehgunaan sistem oleh pengguna adalah elemen penting yang harus dinilai. Aspek ini sangat penting dalam menentukan pendapat pengguna mengenai kebolehgunaan Sistem Sambungan Web bagi Pencegahan Pancingan Data.

Jadual 5.5 Kebolehgunaan

Soalan	Median
Kebolehgunaan	**4.5**
1. Sistem ini dapat mengeluarkan pemberitahuan pop-up tentang laman web yang berunsur pancingan data.	**4.5**
2. Sistem ini dapat menyatakan bahawa laman web tersebut adalah tidak selamat.	**4.5**
3. Sistem ini menyekat laman web yang berunsur pancingan data.	**4.5**
4. Sistem ini dapat membuang penyekatan yang telah dilakukan ke atas sesebuah laman web.	**4.5**
5. Pemaparan maklumat dalam sistem ini adalah tepat dan benar.	**4.5**

d) Kepuasan sistem

Rajah 5.6 menunjukkan faktor yang terakhir ialah kepuasan keseluruhan sistem. Nilai purata keseluruhan merekodkan 4.5. Nilai purata ini menunjukkan pengguna sangat berpuas hati dengan sistem SeaPol ini.

Jadual 5.6 Kepuasan Sistem

Soalan	Median
Kepuasan Sistem	**4.5**
1. Saya berpuas hati dengan sistem ini.	**4.5**
2. Saya berpuas hati dengan antara muka sistem ini.	**4.5**
3. Saya rasa selesa menggunakan sistem ini.	**4.5**
4. Sistem ini berfungsi dengan baiknya.	**4.5**
5. Saya akan cadangkan sistem ini kepada rakan saya.	**4.5**

6 KESIMPULAN

Dalam projek akhir tahun ini, secara keseluiruhannya sebuah sistem sambungan web bagi pencegahan data dipanggil SeaPol telah dibangunkan dengan jayanya. Pembangunan sistem ini dapat memberi manfaat kepada pengguna yang melayari laman web dengan mengesan laman web yang mempunyai unsur pancingan data dan sekaligus menyekatnya daripada berulang di akses kembali. Hal ini dapat memberi kesedaran kepada pengguna tentang kewujudan laman

web berunsur pancingan data melalui Sistem Sambungan Web bagi Pencegahan Pancingan Data.

6.1 Objektif I

Objektif kajian pertama ialah mengenal pasti unsur pancingan data dalam talian. Objektif ini dicapai dengan menetapkan peraturan – peraturan untuk sistem mengesan unsur pancingan data dengan kaedah pengesanan melalui capaian URL dan alamat IP.

6.2 Objektif II

Objektif kajian kedua ialah mereka bentuk sebuah sistem sambungan yang yang mengesan aktiviti pancingan data dalam talian. Objektif ini dicapai dengan membangunkan sebuah sistem sambungan yang bernama SeaPol dengan menggunakan bahasa JavaScript.

6.3 Objektif III

Objektif kajian yang ketiga ialah menilai keberkesanan pendekatan yang direka bentuk dalam mengesan pancingan data. Berdasarkan soal selidik yang telah dijalankan, kebanyakan responden bersetuju bahawa Sistem Sambungan Web bagi Pencegahan Pancingan Data (SeaPol) telah berjaya memainkan peranan sebagai sistem sambungan yang mengesan laman web yang mempunyai unsur pancingan data. Objektif ini berjaya dicapai mengikut analisis soal selidik yang telah dijalankan bersama pengguna.

7 RUJUKAN

Aleroud, A. 2017. Phishing environments, techniques, and countermeasures: A survey. *ScienceDirect,* 24 November: 21.

Alomari, S. 2012. Prevention of Phishing Attacks Based on Discriminative Key Point Features of WebPages. https*://www.researchgate.net/figure/Sequence-Diagram-for-Prevention-of- Phishing-Attack_325568402* [4 Januari 2022].

Bannister, S. 2012. Sharkcop: Google Chrome extension uses machine learning to detect phishing URLs. https*://portswigger.net/daily- swig/sharkcop-google-chrome-extension-uses-machine-learning-to-detect-phishing-urls* [13 Jun 2022].

Barlett, M. t.th. Chrome: Enable/Disable "Not Secure" Warning. https://*www.technipages.com/chrome-enabledisable-not- secure- warning* [26 Januari 2022].

Basset, R. 2019. *5 Common Phishing Techniques*. https*://www.vadesecure.com/en/blog/5-common-phishing-techniques/* [12 Januari 2022].

Chen, H. & Hossain, M. 2021. *Developing a Google Chrome Extension for DetectingPhishing Emails*. https*://easychair.org8* [22 Januari 2022].

Dhamija, R. 2006. Why phishing works. *ACM Digital Library*, 21 Januari: 22.

Ellis, D. t.th. Ways to Recognize a Phishing Email https*://www.securitymetrics.com/blog/ways-recognize-phishing-email* [18 November 2021].

Fruhlinger, J. 2020. What is phishing? How this cyberattack works and how to prevent it. https*://www.csoonline.com/article/2117843/what-is- phishing.html* [18 November 2021].

Groot, N.J. 2021. Phishing Attack Prevention: How to Identify & Avoid Phishing Scams in 2021. http: *//digitalguardian.com/blog/phishing- attack-prevention-how-identify- avoid-phishing-scams* [18 November 2021].

Jinwala, D. 2006. Preventing Phishing Attacks: A Novel Approach. *ResearchGate*, 24 Disember: 21.

Khonji, M. 2013. Phishing Detection: A Literature Survey. *IEEE Xplore*, 24 Disember: 21.

Mahmood, M. 2016. New rule-based phishing detection method. *ScienceDirect*, 28 Februari: 22.

Mark, C.P. 2012. Agile Methodologies and Process Discipline," Crosstalk. *ResearchGate*, 28 Oktober: 21.

Nvisia. 2020. The Agile Process 101 : Understanding the Benefits of Using Agile Methodology. http*://www.nvisia.com/insights/agile-methodology* [28 Oktober 2021].

Porter, K. 2021. What is phishing? How to recognize and avoid phishing scams. http: https://*us.norton.com/internetsecurity-online-scams- what-is- phishing.html* [18 November 2021].

Shah, B. & Dharamishi, K. 2020. Chrome Extension for Detecting Phishing Websites. http: https: *//www.irjet.net/archives/V7/i3/IRJET- V7I3590.pdf* [16 Februari 2021].

Sheng, S. 2016. Fighting against phishing attacks: state of the art and future challenges. *Google Scholar*, 25 May: 22.

Zhang, Y. 2007. Phinding Phish: Evaluating Anti-Phishing Tools. *Google Scholar*, 25 May: 22.

Zaman, N. 2019. Phishing Website Detection: An Improved Accuracy through Feature Selection and Ensemble Learning. *ResearchGate*, 4 May: 22.

Dinie Syahirah Zairol Sahrin (A182536)
Masnizah Mohd
Fakulti Teknologi & Sains Maklumat,
Universiti Kebangsaan Malaysia