



MAESTRÍA EN SOFTWARE PROYECTO DE TITULACIÓN II

SEMANA II: EL MÉTODO CIENTÍFICO-
ANÁLISIS DE LITERATURA
FORMULACIÓN DE LA HIPÓTESIS.

Motivación

2

- **Este capítulo le permitirá:**
 - ▣ Conocer la importancia del Análisis de la Literatura en un investigación;
 - ▣ Comprender cómo se estructura el Marco Teórico y del Estado del Arte;
 - ▣ Aprender cómo se prepara una Revisión sistemática de Literatura;
 - ▣ Comprender cómo se formula una hipótesis de investigación;
 - ▣ Encajar esta información en el protocolo de investigación de su proyecto de titulación.



¿Qué se consigue con el análisis de la literatura y la formulación de la hipótesis?

Contenido

3

- Fundamentación
- El Método científico
 - ▣ Formulación del Problema;
 - ▣ Análisis de la Literatura;
 - ▣ Formulación de la Hipótesis;
 - ▣ Diseño Experimental;
 - ▣ Análisis de Resultados, verificación y validación;
 - ▣ Elaboración del Reporte de Investigación;
- Bibliografía

Análisis de la Literatura

□ Marco Teórico

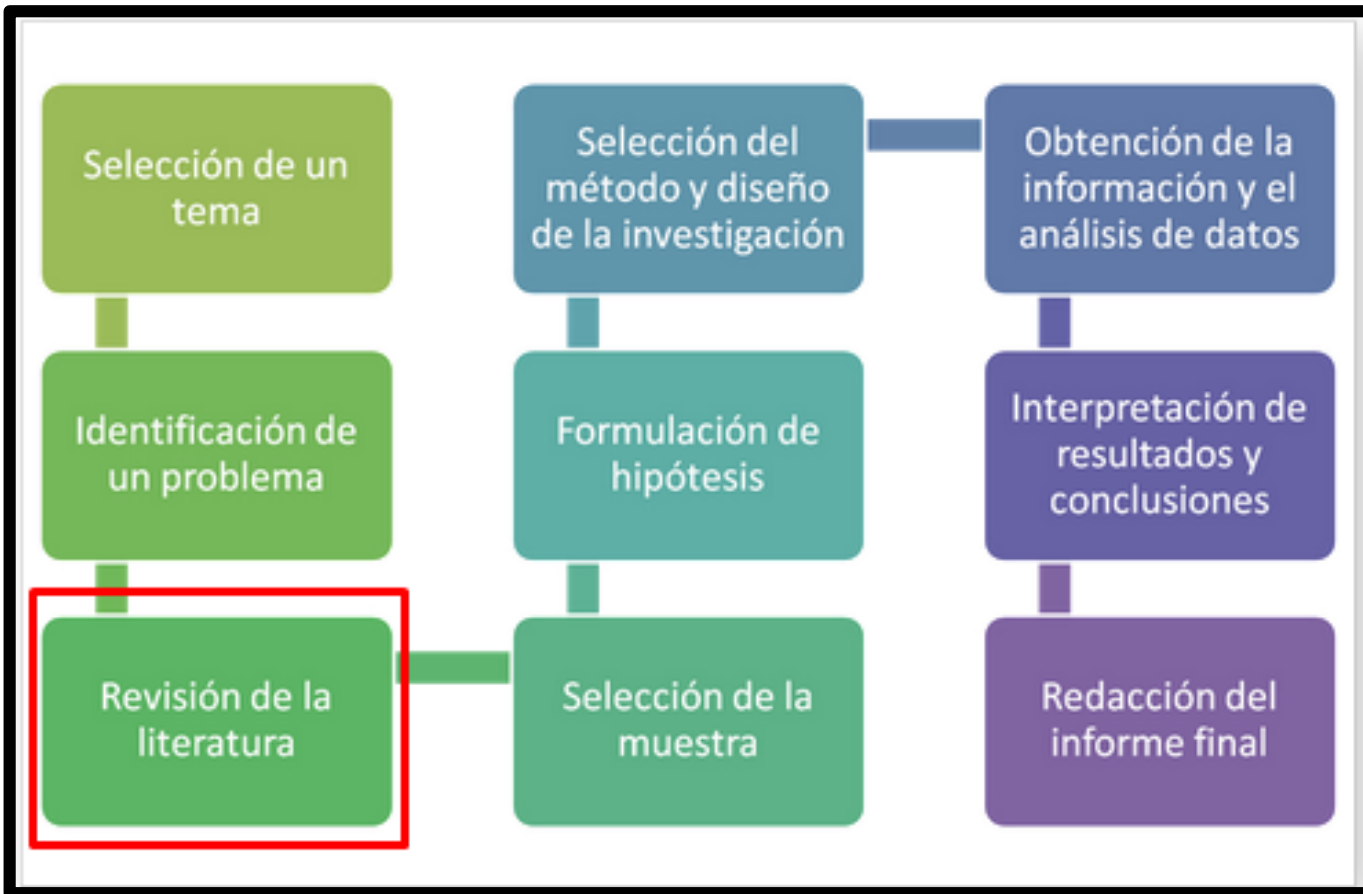
- Es el compendio de una serie de elementos conceptuales que sirven de base a la indagación por realizar;
- Consiste en desarrollar la teoría que va a fundamentar el proyecto con base al planteamiento del problema que se ha realizado.

□ Estado del Arte (Trabajo Relacionados)

- Son los trabajos relevantes de otros investigadores expertos en la misma temática. Establece lo que han investigado otros autores y se incluyen citas de otros proyectos de investigación.



Análisis de la Literatura



Fases de la Investigación e innovación educativa y su relación con el Método Científico

Fuente: Proceso de investigación educativa. Metodología. URL: <https://investigacioneinnovacion4b.wordpress.com/2014/11/02/tema-2-proceso-de-investigacion-educativa-metodologia/>

- *Supone buscar fuentes documentales para construir el marco teórico.
- Exige:
 - ▣ La búsqueda de literatura, que puede ser de tres tipos diferentes.
 - I. *Fuentes primarias*: documentos sobre los que se escribe directamente y proporcionan datos de primera mano;
 - II. *Fuentes secundarias*: procesan la información de las fuentes primarias;
 - III. *Fuentes terciarias*: agrupan compendios de fuentes secundarias.
- Consulta de la literatura.
 - ▣ Leer y extraer datos e ideas necesarias;
 - ▣ Distinguir resumen, contenido y reflexión;
 - ▣ Recoger datos bibliográficos;
 - ▣ Redactar las referencias correctamente.

Análisis de la Literatura

6

□ Marco Teórico

- Una vez que se ha seleccionado el tema objeto de estudio y se han formulado las preguntas que guíen la investigación, el siguiente paso consiste en realizar una **revisión de la literatura** sobre el tema.
- Consiste en **buscar las fuentes bibliográficas** que permitan detectar, extraer y recopilar la información de interés para construir el marco teórico pertinente al problema de investigación planteado.
- Consiste en **desarrollar la teoría que va a fundamentar el proyecto** con base al planteamiento del problema que se ha realizado.
- Aporta con el marco de **referencia conceptual** necesario para delimitar el problema, formular definiciones, fundamentar las hipótesis o las afirmaciones que más tarde tendrán que verificarse, e interpretar los resultados de estudio.
- Consiste en **evitar plagios y repeticiones** de investigaciones



Análisis de la Literatura

7

□ Funciones del Marco Teórico:

- Permite **orientar tanto la investigación** y el **enfoque epistemológico** que se sustenta como la formulación de preguntas, y señala los hechos significativos que deben indagarse.
- Es fundamental porque brinda un **marco de referencia para interpretar los resultados** de la investigación, pues sin teoría es imposible desarrollar una investigación.
- Permite **guiar al investigador para que mantenga su enfoque**, es decir, que este perfectamente centrado en su problema y que impida la desviación del planteamiento original.
- Facilita **establecer afirmaciones** que posteriormente se habrán de someter a la comprobación de la realidad en el trabajo de campo, proceso que ayuda en la inspiración de nuevas líneas y áreas de investigación.

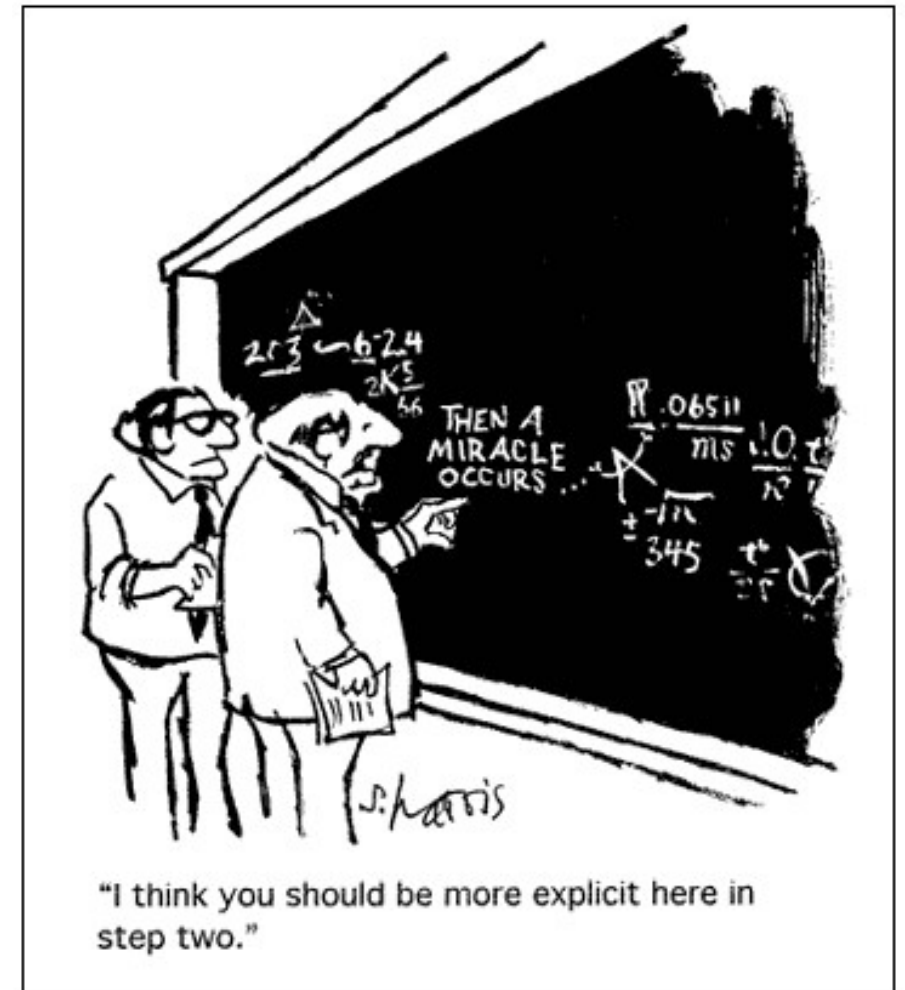


Análisis de la Literatura

8

□ Marco Teórico

- Un buen "marco teórico" no es aquel que contiene muchas páginas, sino el que **trata con profundidad únicamente los aspectos relacionados con el problema**, y vincula lógica y coherentemente los conocimientos, conceptos, variables y proposiciones existentes que se trataran en la investigación.
- La esencia de una tesis de maestría es el conocimiento profundo del tema específico que se ha elegido.
- Es la parte medular de la Tesis y se descompone en:
 - **Antecedentes de la investigación y**
 - **Bases teóricas**



Análisis de la Literatura

□ **Antecedentes de la investigación:**

Es una revisión bibliográfica crítica de:

- ¿Qué tanto se ha investigado sobre el tema?
- ¿Existen estudios previos del que se pretende realizar?
- ¿Quiénes son los estudiosos mas representativos del tema?
- ¿Qué metodología, métodos, técnicas, herramientas han utilizado?
- ¿En que año se ha publicado el estudio y en qué país?
- ¿Cuáles han sido sus primeros hallazgos”

□ **Bases teóricas:** Sirven para entender la realidad y poder explicarla:

- Variable independiente
 - Definición conceptual;
 - Definición Operacional;
 - Métricas, dimensiones, Escalas.
- Variable Dependiente:
 - Definición conceptual;
 - Definición Operacional;
 - Métricas, dimensiones, Escalas;
 - Definición de Instrumentos de Medición
- La influencia de la o las variable Independiente sobre la Dependiente.

Actividad de Aprendizaje

¿En dónde encuentro los antecedentes de investigación?

Actividad de Aprendizaje:

**Tema: Operadores de Búsqueda de Google:
Lista Completa (42 Operadores Avanzados)**

Acceda al URL:

<https://ahrefs.com/blog/es/operadores-de-busqueda-avanzada-de-google/>



Se pide:

- ▣ Acceder al sitio Web;
- ▣ Elegir máximo un operador obligatorio de búsqueda por maestrante y uno al gusto del listado;
- ▣ Realizar ejemplos de búsqueda según sus intereses investigativos y proponerlos a la clase solo en el Google Académico; :
 - Antecedentes de la Investigación;
 - Bases teóricas.
- ▣ Cada alumno tiene que explicar el procedimiento y mostrarlo en clase;
- ▣ Guardar en una “bitácora”, con una estructura cronológica que se debe actualizar regularmente sobre un tema concreto: que contenga al menos 10 publicaciones primarias de cada aspecto, con el autor, título, el año, país, Tipo (Congreso, revista, libro), Database. y la referencia bibliográfica en formato APA.

Revisión Sistemática de Literatura

- Una Revisión Sistemática de la Literatura (Systematic Literature Review, SLR) es un tipo de revisión de la literatura que recopila y analiza críticamente múltiples estudios o trabajos de investigación a través de un **proceso sistemático**
- Es un método sistemático para identificar, evaluar e interpretar el trabajo de investigadores, académicos y profesionales en un campo elegido (Fink, 1998).
- El objetivo de una RSL es proporcionar un resumen exhaustivo de la literatura disponible pertinente a las **preguntas de investigación.**

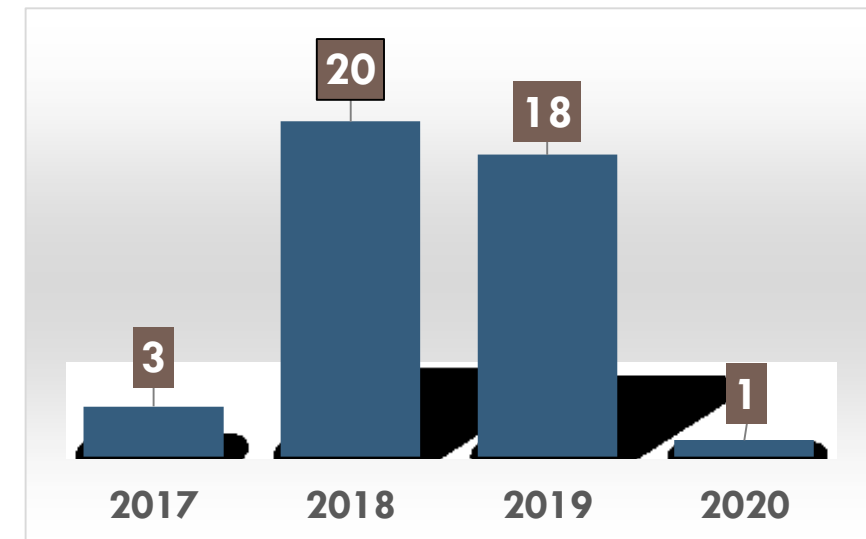


FIGURE 8. Papers by Year.

Revisión Sistemática de Literatura

□ Propósitos:

- Alcanzar un conocimiento más profundo en su campo del conocimiento;
- Sentar las bases del estado del arte de un trabajo de investigación;
- Obtener información sobre las tendencias actuales y los desafíos futuros;
- Identificar los autores más importantes;
- Identificar las revistas y conferencias más importantes;
- Caracterizar la forma de solución de las publicaciones;
- Obtener citas. (85)(150*).

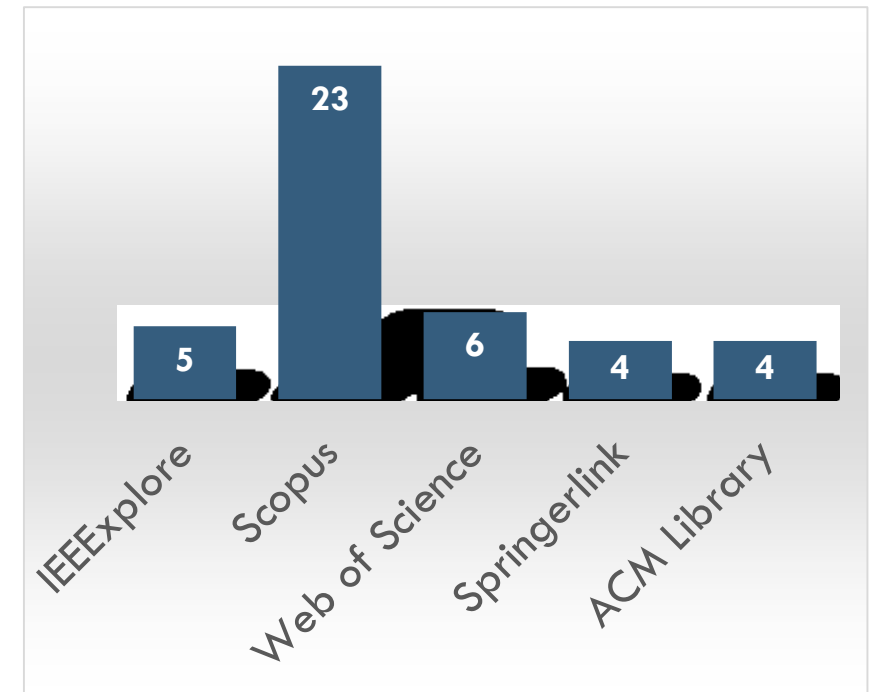


FIGURE 5. Number of papers by Scientific Database

Revisión Sistemática de Literatura - Protocolo

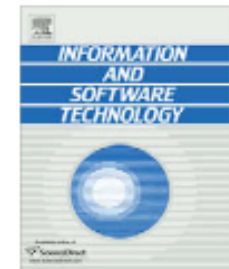
Information and Software Technology 51 (2009) 7–15



Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof



Systematic literature reviews in software engineering – A systematic literature review

Barbara Kitchenham^{a,*}, O. Pearl Brereton^a, David Budgen^b, Mark Turner^a, John Bailey^b, Stephen Linkman^a

^aSoftware Engineering Group, School of Computer Science and Mathematics, Keele University, Keele Village, Keele, Staffs, ST5 5BG, UK

^bDepartment of Computer Science, Durham University, Durham, UK

ARTICLE INFO

ABSTRACT

Revisión Sistemática de Literatura - Protocolo

Contents

1.	Introduction	8
2.	Method	8
2.1.	Research questions	8
2.2.	Search process	8
2.3.	Inclusion and exclusion criteria	9
2.4.	Quality assessment	9
2.5.	Data collection	9
2.6.	Data analysis	9
2.7.	Deviations from protocol	9
3.	Results	10
3.1.	Search results	10
3.2.	Quality evaluation of SLRs	10
3.3.	Quality factors	10
4.	Discussion	11
4.1.	How much EBSE Activity has there been since 2004?	11
4.2.	What research topics are being addressed?	11
4.3.	Who is leading EBSE research?	12
4.4.	What are the limitations of current research?	12
4.5.	Limitations of this study	13
5.	Conclusions	13
	Acknowledgements	15
	References	15

Revisión Sistemática de Literatura

□ **Protocolo:**

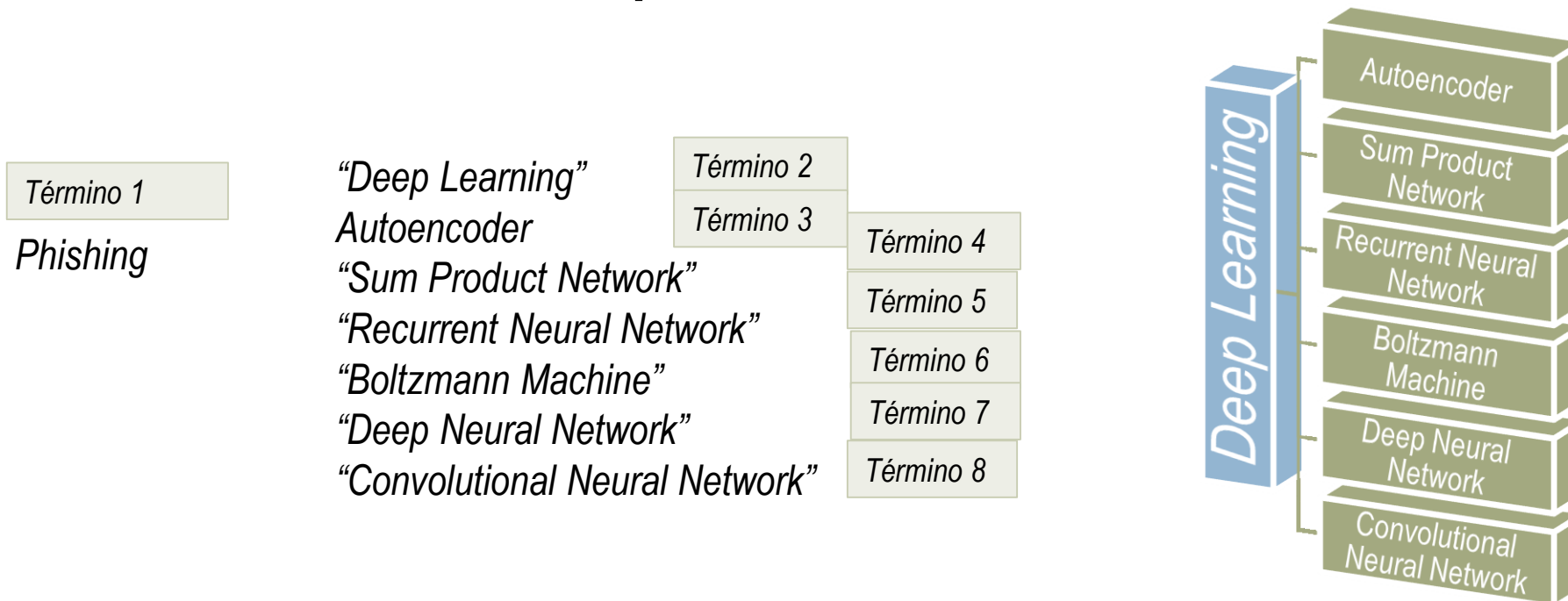
1. Definir las preguntas de investigación;
2. Definir criterios de inclusión y exclusión para la RSL;
3. Identificar las bases de datos y motores de búsqueda que se van a utilizar;
4. Definir los términos de búsqueda;
5. Buscar en bases de datos científicas;
6. Fases de Revisión;
7. Extracción de Datos;
8. Presentación de Resultados.

Revisión Sistemática de Literatura- Ejemplo

1. **Definir la pregunta de investigación** (Research Question – RQ1).
 - ▣ Tema de Investigación: Una nueva Técnica de **Deep Learning** para **Detectar y Mitigar** los Ataques de **Phishing**
 - ▣ **RQ1**: ¿Que técnicas de **Deep Learning** se han aplicado actualmente para **detectar y mitigar** Ataques de **Phishing**?
2. **Definir criterios de inclusión y exclusión**
 - ▣ Solo estudios primarios. (*La RSL no es un estudio primario);
 - ▣ Solo investigaciones con resultados. Porcentaje, exactitud, etc.;
 - ▣ Escritos solamente en Ingles;
 - ▣ Estudios de los últimos 5 años;
 - ▣ Describe la aplicación de al menos una técnica de Deep Learning en la detección de ataques de Phishing;
 - ▣ Temas principales: Phishing AND Deep Learning;
3. **Identificar las bases de datos y motores de búsqueda**
 - ▣ Springer; IEEE Xplorer, Web of Science; Science Direct; ACM; Taylor and Francis;Scopus

Revisión Sistemática de Literatura - Ejemplo

4. Definir los términos de búsqueda



Cadena de Búsqueda:

- Phishing and (*"Deep Learning"* or Autoencoder or *"Sum Product Network"* or *"Recurrent Neural Network"* or *"Boltzmann Machine"* or *"Deep Neural Network"* or *"Convolutional Neural Network"*)

Revisión Sistemática de Literatura - Ejemplo

5. Buscar en bases de datos científicas y extraer contenidos y datos relevantes

Buscar en Metadatos:

- Título
- Abstract o Resumen
- Palabras Claves.

No metadatos:

- Introducción
- Conclusiones
- Cuerpo

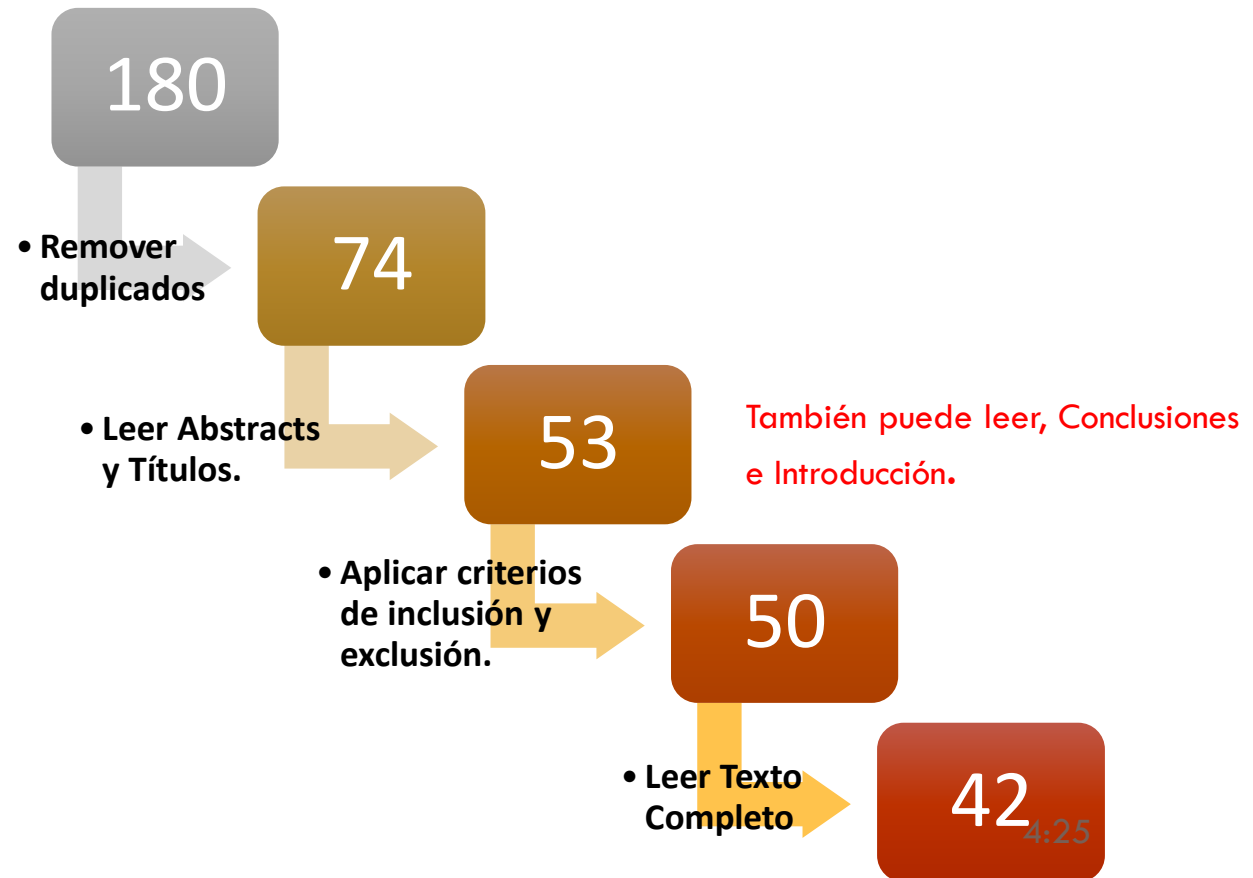
The screenshot shows the IEEE Xplore Advanced Search interface. The top navigation bar includes 'IEEE Xplore', 'Browse', 'My Settings', 'Help', and 'Institutional Sign In'. Below this is a search bar with a dropdown menu set to 'All' and a search button. The main section is titled 'Advanced Search' and has tabs for 'Advanced Search', 'Command Search', and 'Citation Search'. The 'Advanced Search' tab is selected. Below the tabs, there is a section titled 'Enter keywords, select fields, and select operators'. This section contains a table of search terms and operators. The search terms are: 'Phishing', 'Deep Learning', 'Autoencoder', 'Sum Product Network', 'Recurrent Neural Network', 'Boltzmann Machine', 'Deep Neural Network', and 'Convolutional Neural Network'. The operators are: 'AND', 'OR', 'OR', 'OR', 'OR', 'OR', 'OR', and 'OR'. The search fields are all set to 'All Metadata'. To the right of the search terms are buttons for 'Learn More', 'Data Fields', 'Search Exan', 'Search Oper', and 'Search Tips'.

Search Term	Operator	Search Term	Operator	Search Term	Operator	Search Term	Operator	Search Term	Operator	Search Term	Operator	Search Term	Operator	Search Term	Operator	Search Term	Operator
Phishing																	
AND		Deep Learning															
OR		Autoencoder															
OR		Sum Product Network															
OR		Recurrent Neural Network															
OR		Boltzmann Machine															
OR		Deep Neural Network															
OR		Convolutional Neural Network															

Revisión Sistemática de Literatura - Ejemplo

6. Fase de Revisión

IEEE Explore	31
Scopus	80
ACM Library	20
Web of Science	10
Taylor & Francis	2
Google Scholar	37
	<hr/>
	180



Revisión Sistemática de Literatura - Ejemplo

7. Extracción de Datos

Database	Paper		ABSTRACT				WEBPAGE						EMAIL											OTHER														What they did?	Conclusions	Accuracy
			AE	SPN	RNN	BM	DNN	CNN	SAE	DAE	SDAE	LSTM	ESN	DBM	RMB	DBN	GAN	DRBM	ReNN																					
Scopus	Evaluating deep lea	Yes	1		1			1			1					1											Compæ LSTM anc	99.97%												
Scopus	Web phishing detec	Yes	1		1				1																1		This paper propc	90% TPI												
Springerlink	Personalized: Brow	Yes			1						1																In this work we i	99%												
Springerlink	Phishing Analysis of	Yes	1		1							1															In this Se realizc	84%												
Springerlink	Phishing Detection I	Yes	1																						1		Furthe Se utilizai	96.50%												
Springerlink	Phishing Detection I	Yes	1					1							1												This paper design	99.14%												
Springerlink	Detecting malicious	Yes	1													1											We err which ca	98.52%												
Taylor & Fran	Deep belief network	Yes	1						1	1															1		The resul	89.00%												
Web of Scien	A Deep-Learning-Dr	Yes	1							1	1																In this paper [9],	86.63%												
Web of Scien	PDRCNN: Precise PI	Yes	1					1			1				1														97%											
Web of Scien	Ensemble one-vs-	Yes		1							1																		95.50%											
Scopus	Classification of Phi	Yes																																						
Scopus	Comparison of Ense	Yes		1																							Two types of NN	94.41%												
Scopus	Hybrid intelligent pt	Yes	1							1																														
Scopus	Parameter Setting f	Yes	1																								In this pa	96.65%												
Scopus	Deep learning baser	Yes	1	1						1																	In X, the authors	92%												
Scopus	A Survey on Deep Li	Yes																																						
Scopus	Phishing URL detect	Yes	1					1			1																			97.90%										
Scopus	Phishing URL detect	Yes	1					1			1																			96.48%										
►	Solo Soluciones 24 feb		Paper by Data				Paper by SubTechnique Applied						Paper by Technique Applied						Paper type			References by Paper																		

Revisión Sistemática de Literatura - Ejemplo

8. Presentación de Resultados

TABLE I
SELECTED PRIMARY STUDIES

NO.	TITLED	REF.	DATABASE	PAPER TYPE	SJR QUARTILE	TOTAL REFS.	YEAR	COUNTRY
1	Phishing Email Detection Using Improved RCNN Model with Multilevel Vectors and Attention Mechanism	[15]	Scopus	Journal	Q1	51	2019	China
2	Phishing Website Detection Based on Multidimensional Features Driven by deep learning	[29]	Web of Science	Journal	Q1	38	2019	China
3	Evaluating deep learning approaches to characterize and classify malicious URL's	[30]	Web of Science	Journal	Q2	27	2018	India
4	Personalized, Browser-Based Visual Phishing Detection Based on deep learning	[31]	IEEEExplore	Book Series	Q2	6	2019	Italy
5	Phishing Detection Method Based on Borderline-Smote Deep Belief Network	[32]	IEEEExplore	Book Series	Q2	11	2017	China
6	PDRCNN: Precise Phishing Detection with Recurrent Convolutional Neural Networks	[33]	Web of Science	Journal	Q2	34	2019	China
7	Ensemble one-vs-all learning technique with emphatic & rehearsal training for phishing email classification using psychology	[34]	Web of Science	Journal	Q2	60	2018	United States
8	Hybrid intelligent phishing website prediction using deep neural networks with genetic algorithm-based feature selection and weighting	[35]	Scopus	Journal	Q2	50	2019	Saudi Arabia
9	A deep learning model with hierarchical LSTMs and supervised attention for anti-phishing	[36]	Springerlink	Conference Paper	Q3	39	2018	Viet Nam, United States
10	A platform for automatic identification of phishing URLs in mobile text messages	[37]	Springerlink	Journal	Q3	15	2018	China
11	A.R.E.S: Automatic rogue email spotter crypt coyotes	[38]	Springerlink	Journal	Q3	35	2018	India
12	CNN-based malicious user detection in social networks	[39]	Springerlink	Journal	Q3	23	2018	South Korea
13	Deep learning-based phishing E-mail detection CEN-Deepspam	[9]	Scopus	Journal	Q3	20	2018	India

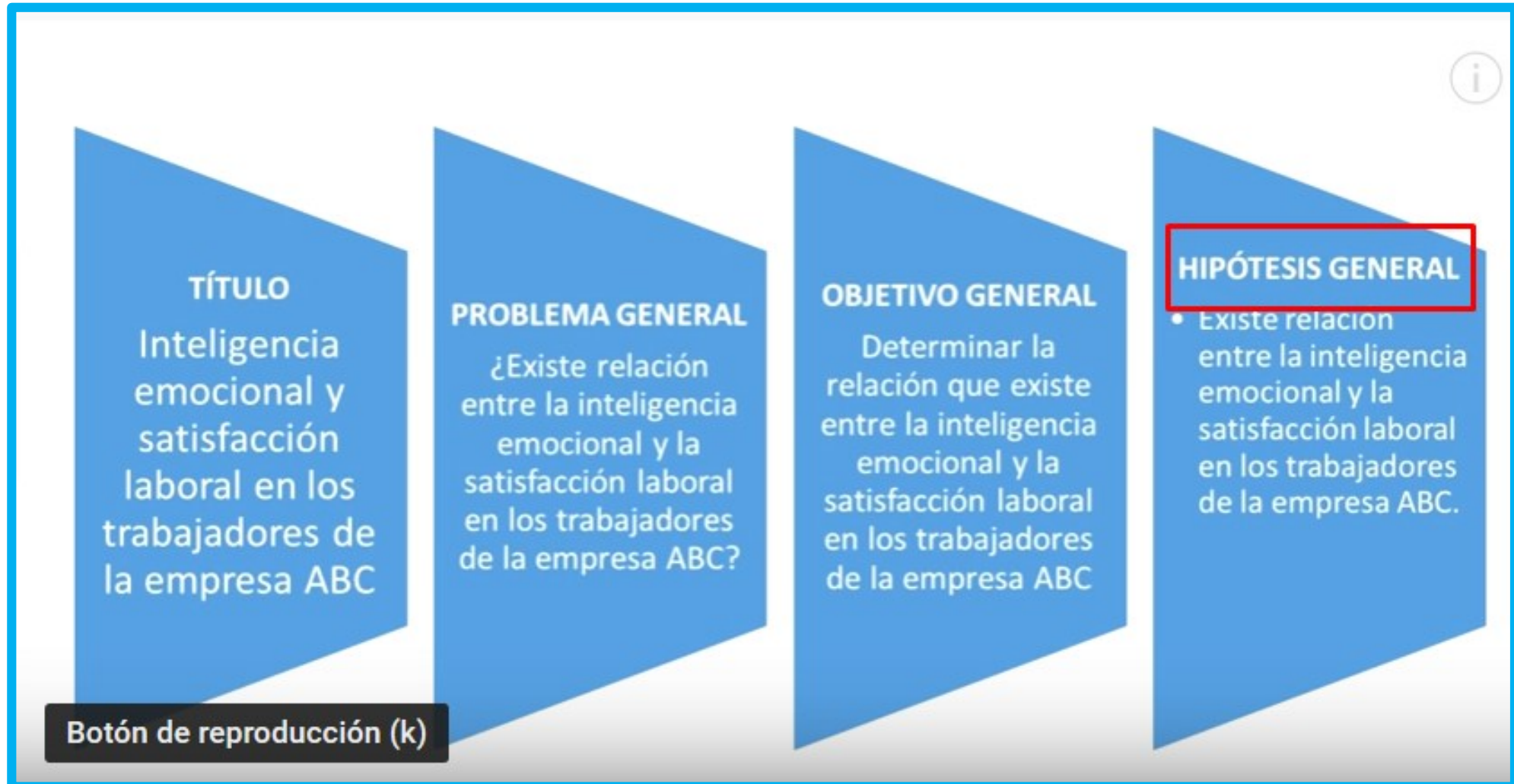
Fuente: Revisión Sistemática de la Literatura (SLR). De la Teoría a la Práctica, Eduardo Benavides, julio 2020. E-mail: debenavides@espe.edu.ec

Formulación de la Hipótesis

- **Formulación de la Hipótesis**
 - Es la manera o los requisitos que deberán cumplirse para proceder a redactarla;
 - El enunciado escrito es conocido como hipótesis;
 - Es un supuesto, una proposición o una oración que va a ser contrastada y que puede ser V o F;
 - Es una respuesta anticipada o tentativa a un planteamiento del problema;
 - Tiene relación directa con la pregunta principal de investigación.



Formulación de la Hipótesis



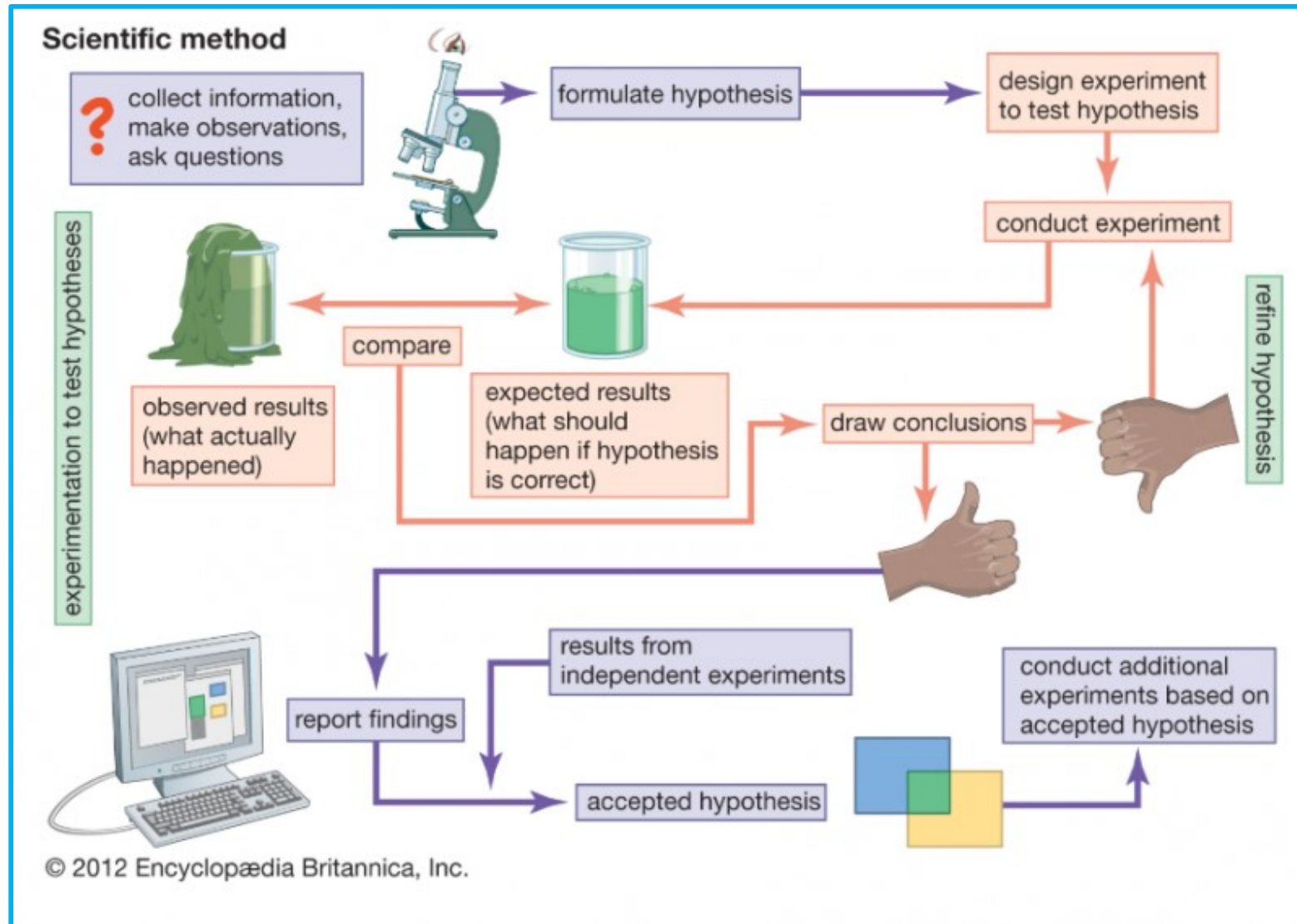
Formulación de la Hipótesis

- Una hipótesis de investigación es una **proposición** o **declaración predictiva específica, clara y comprobable** sobre el posible resultado de un estudio de investigación científica basado en una propiedad particular de una población, como presuntas diferencias entre grupos en una variable particular o relación entre variables.
- Especificar las hipótesis de investigación es uno de los pasos más importantes en la **planificación** de un estudio de investigación científica cuantitativa.



Fuente: Biofilia hipótesis diseño gráfico, interpretación, texto, logo.png. Disponible en: URL <https://www.pngegg.com/es/png-hqeij>

Formulación de la Hipótesis

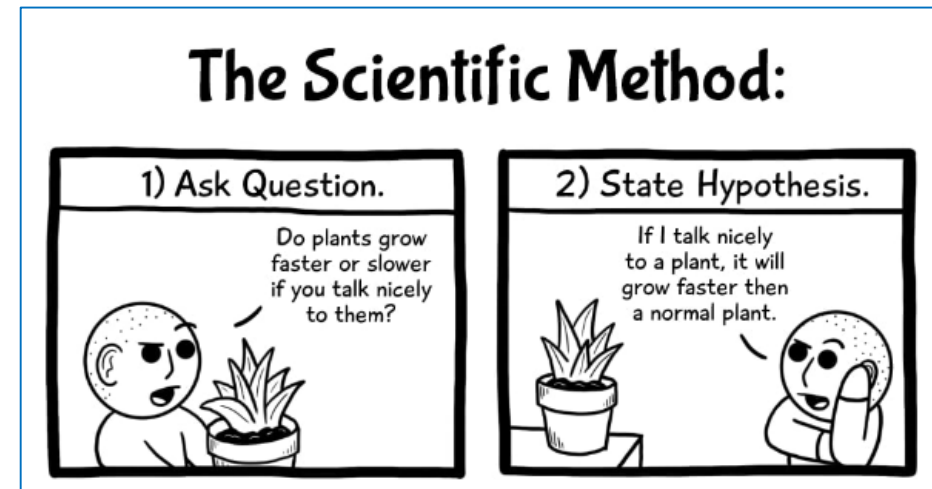


Formulación de la Hipótesis

26

□ Hipótesis de Investigación

- Se define como una aseveración, conjetura o proposición sobre las probables relaciones entre dos o mas variables;
- Con frecuencia se pueden expresar en forma descriptiva, correlacional, de causalidad, de nulidad, etc., dependiendo del propósito y naturaleza de la investigación que se intenta desarrollar;
- Normalmente se formulan en investigaciones cuantitativas y en menor frecuencia en las cualitativas.



Las Hipótesis sirven de:

**Guía o criterio de orientación de la investigación;
Una vez probadas se convierten en tesis y son puntos de partida para
nuevas investigaciones.**

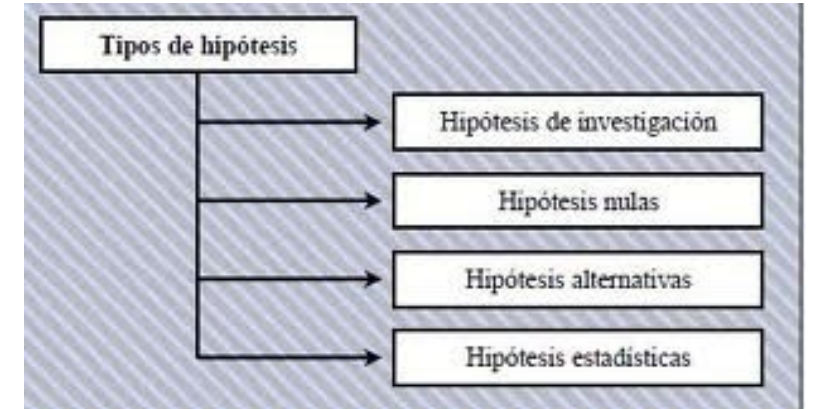
**Fuente directa para la operacionalización de las variables y sus
respectivos indicadores.**

Formulación de la Hipótesis

27

□ Tipos de Hipótesis:

- ▣ **Hipótesis de investigación** es la hipótesis que se desea comprobar:
 - **Hipótesis descriptivas** (Estudios con una sola variable);
 - **Hipótesis correlacionales** (Correlación de dos variables);
 - **Hipótesis de Causalidad** (Causa – efecto de dos variables).
- ▣ **La hipótesis nula:** es aquella que se plantea para ser rechazadas al realizar una prueba de hipótesis para inferir características de una población. Sirven para refutar o negar lo que afirma la hipótesis de investigación.
- ▣ **La hipótesis alternativa** es igualmente una afirmación acerca de la población de origen. Muchas veces, es una hipótesis alterna a la H_i .



♦Cuál es la probabilidad de que ocurra la hipótesis nula?



Las pruebas de significancia estadística miden esta probabilidad

Formulación de la Hipótesis

28

□ Variables en una Hipótesis

- ▣ **Variable dependiente.** En un estudio experimental, la variable dependiente es la característica que se investiga y que siempre debe ser evaluada o manipulada.
- ▣ **Variable independiente.** Es la característica que se puede medir por separado y que puede ser causa de la variable dependiente.

El objetivo de un **experimento es comprobar** cuando una o mas variables independientes influyen o alteran a la variable dependiente. Es decir, si la variable independiente produce algún cambio significativo en la variable dependiente. Esta actividad requiere de controlar estrictamente las condiciones experimentales de un estudio.

Formulación de la Hipótesis

29

- **Estructura de la hipótesis:** Es una suposición en la que se relacionan dos o más variables. Consta de:
 - ▣ **Unidad de observación** (sobre qué versa la investigación), un país, un colectivo, acontecimientos, grupos, etc.;
 - ▣ **Variables** (Aspectos cuantitativos y cualitativos que son objetos de búsqueda ante las unidades de observación);
 - ▣ **Términos lógicos o relacionales** (relaciones de las unidades de observación con las variables o entre estas últimas entre sí).
 - ▣ Ejemplo:
 - **Si** no existe estudio individual, **entonces** habrá bajas calificaciones.
 - **Unidad de observación:** Los estudiantes
 - **Variables:** Estudio individual/bajas calificaciones
 - **Problema:** Bajo rendimiento escolar
 - **Términos lógicos:** relación si/entonces

Formulación de la Hipótesis-Tipos

30

- **Hipótesis Descriptiva:** son aquellas que involucran una sola variable. Se utiliza en investigaciones de tipo descriptivo, como pudieran ser los estudios por encuesta.
 - **Ejemplos:**
 - El periodo de recuperación de la inversión del proyecto de *investigación de sistemas distribuidos* es de dos años;
 - Los productos de consumo doméstico en Ecuador aumentarán un 18 % en los próximos seis meses;
 - La expectativa del ingreso mensual de los docentes en el Ecuador para el 2021 es de 2400 USD anuales.

Formulación de la Hipótesis

31

- **Hipótesis Correlacional:** Expresa una posible asociación o relación entre dos o mas variables, sin que sea importante el orden de presentación de las variables, ya que no expresan una relación de causalidad. Para verificarlas se utilizan pruebas estadísticas de correlación.

- **Ejemplos:**

- A mayor apreciación del dólar norteamericano, mayor depreciación del peso mexicano;
- El volumen de importaciones en Chile disminuye con el aumento en el tipo de cambio peso-dólar;
- La inteligencia está relacionada con la memoria;
- A mayor motivación por el estudio, mayor manifestación por aprender;
- A mayor número de hijos menor nivel de vida tiene la familia en América Latina;
- Existe relación entre la motivación y la productividad.

Formulación de la Hipótesis

32

- **Hipótesis de Causalidad.-** Las hipótesis de causalidad se formulan para investigaciones experimentales. Expresan una relación de causa-efecto entre las variables que se someten a estudio (Variable Independiente e Independiente);
- Puede expresar una relación causal entre una variable independiente y una variable dependiente, o bien, puede hacerlo entre mas de una variable independiente y una variable dependiente.
 - ▣ Ejemplos:
 - El elevado índice de inflación en México es **causa** del bajo poder adquisitivo del peso mexicano;
 - Los factores de productividad total (insumo humano, materia prima, energía, capital y otros gastos) del sector manufacturero mexicano son los **determinantes** de la productividad total;
 - La desintegración familiar **provoca (causa)** el bajo rendimiento académico de los hijos;

Formulación de la Hipótesis

1.3. Hipótesis de Causalidad

- Se caracterizan porque presentan la V.I. y V.D.
- Se probará la relación Causa_Efecto.

Ejemplo:

- La disintegración familiar provoca el bajo autoestima.
V. I. V.D.
- El rendimiento academico influye en la deserción escolar.
V. I. V.D.

Formulación de la Hipótesis - EJEMPLO

34

□ Operacionalización de Variables (Ejemplo)

▣ Hipótesis de Investigación:

■ **Si** se implementa un mecanismo de gestión del overhead en una infraestructura virtualizada, **entonces** se puede incrementar su desempeño.

■ Variables Independientes:

- Penalización (overhead) un entorno virtualizado en ambiente Linux. La variable está definida como x1. Se mide mediante el % de Consumo del CPU.
- Capacidad de Transmisión. La Variable está definida como x2 y mide el ancho de banda y conectividad para acceso a usuarios, mediante las métricas de desempeño de redes IP (IPPM –RFC 4148).
- Número de Máquinas virtuales. La variable está definida como x3 y mide el número de equipos virtualizados.

■ Variable Dependiente

- Desempeño en un entorno virtualizado. Depende de las variables independientes (x1, x2, x3). Esta variable estará definida como y cuyos valores discretos estarán dados por: Se incrementa, No se incrementa:
- $Y=f(x1+x2+x3);$

Formulación de la Hipótesis -EJEMPLO

35

□ Matriz de Operacionalización de Variables

Variable	(Definición Conceptual)	Métricas, Dimensiones (Definición Operacional)	indicador
Overhead (Independiente)	Es la penalización producida por la capa de virtualización. Es el porcentaje del tiempo del CPU y la asignación a la supervisión del mismo y no a los usuarios	% tiempo de consumo de CPU % del consumo del CPU Overlap	% del tiempo % CPU Sumatoria % de CPUs
Capacidad de Transmisión (Independiente)	La capacidad de un canal es la cantidad de información (en bits) que se puede transmitir a través de un canal durante 1 segundo.	C(tx) Ancho de Banda latencia	b/s Hertzios Ms
Desempeño en un Entorno Virtualizado (Dependiente)	Es la sumatoria del tiempo de CPU ocupado por los programas de c/MV sobre el tiempo total transcurrido.	Rendimiento Throughput Tiempo de respuesta

Actividad de Aprendizaje

¿Cómo se formula una hipótesis de investigación?

¿Cómo se definen y se miden las variables?

Actividad de Aprendizaje:

Tema: Formulación de la Hipótesis de su trabajo de titulación y Definición conceptual y operacional de Variables

$$y = f(x)$$

↑
Variable dependiente

↑
Variable independiente

Se pide:

- Formular la hipótesis de su trabajo de titulación, siguiendo los siguientes pasos:
 - Identificar el problema
 - Identificar la variable dependiente y (el efecto) (lo que se conoce)
 - Identificar la variable independiente x (las causas) (lo que no se conoce)
 - Escribir una suposición del porque ocurre y
 - Redactar en términos de **Si** X **entonces** Y
- Realizar la Definición Operacional de las variables de la hipótesis de investigación de su proyecto de titulación, utilizando la matriz ejemplo;
- Cada maestrante debe explicar en clase, ¿cuál es su hipótesis general? ¿cuál es la variable que se va a manipular?, ¿cuál no se puede manipular?, ¿cómo se medirán las dos?, ¿cuáles son sus métricas, e indicadores?

Actividad de Aprendizaje

$y = f(x)$

↑
Variable dependiente

↑
Variable independiente

Variable	(Definición Conceptual)	Métricas, Dimensiones (Definición Operacional)	indicador
(Independiente)			
(Dependiente)			
Otras			



HP Software University Association (HPSUA)
14th Workshop Garching/Munich, Germany
8th – 11th July 2007

A quantitative comparison of virtual network environments based on performance measurements

W. M. Fuertes

Escuela Politécnica Superior – Universidad Autónoma de Madrid, Madrid, Spain.
walter.fuertes@estudiante.uam.es

J. E. López de Vergara

Escuela Politécnica Superior – Universidad Autónoma de Madrid, Madrid, Spain
jorge.lopez_vergara@uam.es

Aim and objectives

- Virtualization tools can be an alternative to implement networking scenarios, with the purpose of making measurement experiments similar to those in real networks. For this, it is important to use the virtualization tool that consumes fewer resources.
- In this work, we have obtained a quantitative comparison of the CPU and memory consumption of those virtualization tools during the boot up and execution of a concrete virtual scenario.

A quantitative comparison of virtual network environments based on performance measurements

Walter M. Fuertes and Jorge E. López de Vergara

Dept. Ingeniería Informática, Escuela Politécnica Superior,
Av. Francisco Tomás y Valiente, 11,
Universidad Autónoma de Madrid, Madrid, Spain.
walter.fuertes@estudiante.uam.es, jorge.lopez_vergara@uam.es

Abstract. Virtualization tools can be an alternative to implement networking scenarios, with the purpose of making measurement experiments similar to those in real networks. For this, it is important to use the virtualization tool that consumes fewer resources. In this work, we have obtained a quantitative comparison of the CPU and memory consumption of these virtualization tools during the boot up and execution of a concrete virtual scenario.

Keywords: Virtualization tools, virtual machine, performance evaluation.

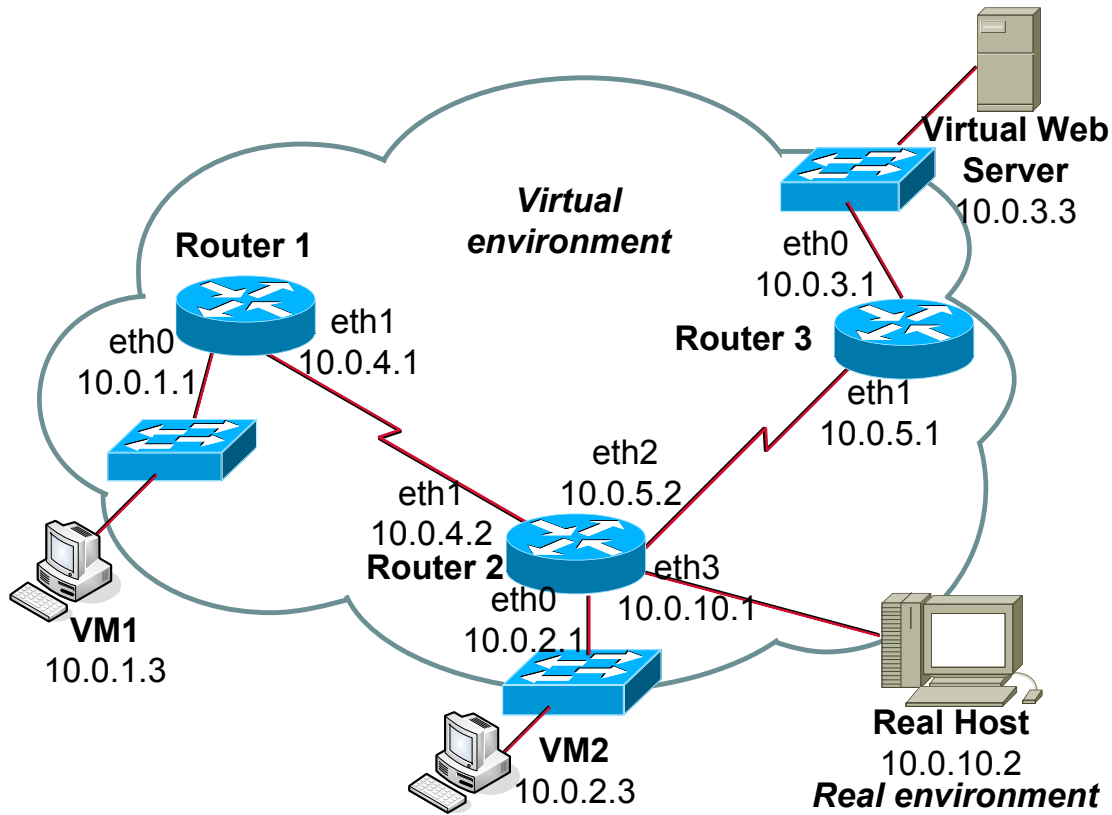
1. Introduction

At present, products and services have to be deployed in networks. However, these networks have other services already in use, and the deployment of new services can cause congestion to the networks, degrading some services and reducing their performance. At the same time, users require that these services assure a certain quality level. Nevertheless, it is uncertain what service quality can be provided a priori on these networks. Then, it is necessary to rely on an infrastructure that estimates the services' behavior, as well as measurement procedures and tools to make tests and to measure the delivered quality of these services.

As alternative we suggest the use of virtual computing platforms, reproducing real infrastructures in a virtual environment. This reduces the risk of damaging real networks, as well as the cost of the development and experimentation. At the same time, the service functionality can be tested, and the results obtained are going to be close to what should be measured in a real scenario, given that it is the real application, and not a simulated model, what is being executed. Even more, as shown in [1], virtualization results are very similar to simulation results. Within this context, the fundamental problem consists of identifying what virtualization tool consume less CPU and memory and is most suitable to make experiments related with the networking.

This research work is intended to verify the functionality of real network situations within virtual environments using a single computer. The result allows making diverse experiments in a virtual network platform to measure the performance and quality of deployed services. These experiments will provide valuable data to calculate the

Design And Implementation



- Clocks synchronized with an NTP server
- Installation and configuration of each virtualization tool
 - ▣ We tried to use the same file system and similar kernel on all the tools for a fair comparison.
 - ▣ Later we cloned the first VM to the remaining VMs, in order to reduce installation time.
 - ▣ We added virtual interfaces, configured IP address and started services.
- Creation and execution of programs that automatically deployed and started all scenarios.
- Implementation of an algorithm, which allows us the measurement of the CPU and Memory consumption, during the boot up and execution.
 - ▣ The execution consisted on a file download from the virtual web server to the real host.

Results

Networking scenarios deployed with VNUML and XEN

```
Virtual Console #0
HTTP request sent, awaiting response... 200 OK
Length: 808,880 (790K) [text/html]
100%[=====] 808,880  --K/s

11:34:01 (14.12 MB/s) = 'wal.html' saved [808880/808880]

uml1:~# ls -l
total 788
drwxr-xr-x 3 root root 1024 Aug 2 2006 src
-rw-r--r-- 1 root root 808880 Feb 23 11:31 wal.html
uml1:~# ls -l
total 788
drwxr-xr-x 3 root root 1024 Aug 2 2006 src
-rw-r--r-- 1 root root 808880 Feb 23 11:31 wal.html
uml1:~# ping 10.0.3.3
PING 10.0.3.3 (10.0.3.3) 56(84) bytes of data:
64 bytes from 10.0.3.3: icmp_seq=1 ttl=61 time=83.6 ms
64 bytes from 10.0.3.3: icmp_seq=2 ttl=61 time=0.446 ms
--- 10.0.3.3 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 101ms
rtt min/avg/max/mdev = 0.446/32.026/63.607/31.581 ms
uml1:~#
```

```
Virtual Console #0
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
uml2:~# ifconfig
eth2
Link encap:Ethernet HWaddr FE:FD:00:00:02:02
inet addr:10.0.2.3 Bcast:10.0.2.255 Mask:255.255.255.0
inet6 addr: fe80::fcdfffffe0020202/64 ScopeLink
UP BROADCAST RUNNING PROMISC MULTICAST MTU:1500 Metric:1
RX packets:14 errors:0 dropped:0 overruns:0 frame:0
TX packets:6 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:1000
RX bytes:352 (352.0 b) TX bytes:468 (468.0 b)
Interrupt:5

Size Used Avail Use% Mounted on
727M 571M 134M 82% /
46M 0 14M 0% /dev/shm
354K 354K 0 100% /mnt/umlal

uml2:~#
```

```
Virtual Console #0
collisions:0 txqueuelen:1000
RX bytes:301397198 (287.4 MiB) TX bytes:7925202 (7.5 MiB)
Interrupt:169

lug:/home/walter# ping -c 3 10.0.10.1
PING 10.0.10.1 (10.0.10.1) 56(84) bytes of data:
64 bytes from 10.0.10.1: icmp_seq=1 ttl=64 time=0.1 ms
64 bytes from 10.0.10.1: icmp_seq=2 ttl=64 time=0.7 ms
64 bytes from 10.0.10.1: icmp_seq=3 ttl=64 time=0.6 ms
--- 10.0.10.1 ping statistics ---
3 packets transmitted, 3 received, 0% packet loss, time 100ms
rtt min/avg/max/mdev = 0.1/0.2/0.7/0.2 ms
lug:~#
```

```
Virtual Console #0
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/*copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
uml4:~# telnet zebra
telnet: could not resolve zebra/telnet: Temporary failure in name resolution
uml4:~# /etc/init.d/zebra stop
Stopping Zebra daemons: (zebra) (bgpd) (ripd) (ripngd) (ospfd) (ospf6d) (zebra).
uml4:~# /etc/init.d/quagga start
Loading capability module if not yet done.
Starting Quagga daemons (priority):
uml4:~# man quagga
No manual entry for quagga
uml4:~# telnet quagga
telnet: could not resolve quagga/telnet: Temporary failure in name resolution
uml4:~# /etc/init.d/quagga stop
Stopping Quagga daemons (priority): (zebra) (bgpd) (ripd) (ripngd) (ospfd) (ospf6d)
(isisd).
Removing all routes made by zebra.
Nothing to flush.
uml4:~#
```

```
Virtual Console #0
uml16:~# ifconfig eth1
eth1
Link encap:Ethernet HWaddr FE:FD:00:00:06:01
inet addr:10.0.5.1 Bcast:10.0.5.255 Mask:255.255.255.0
inet6 addr: fe80::fcdfffffe0060101/64 ScopeLink
UP BROADCAST RUNNING PROMISC MULTICAST MTU:1500 Metric:1
RX packets:20 errors:0 dropped:0 overruns:0 frame:0
TX packets:3 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:1000
RX bytes:0 (0.0 b) TX bytes:338 (338.0 b)
Interrupt:5

uml16:~# ifconfig eth2
eth2
Link encap:Ethernet HWaddr FE:FD:00:00:06:02
inet addr:10.0.5.1 Bcast:10.0.5.255 Mask:255.255.255.0
inet6 addr: fe80::fcdfffffe0060202/64 ScopeLink
UP BROADCAST RUNNING PROMISC MULTICAST MTU:1500 Metric:1
RX packets:20 errors:0 dropped:0 overruns:0 frame:0
TX packets:3 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:1000
RX bytes:0 (0.0 b) TX bytes:338 (338.0 b)
Interrupt:5
```

```
Xen1
Linux (none) 2.6.18-4-xen-686 #1 SMP Mon Mar 26 21:49:04 UTC 2007 1686 GNU/Linux
(none):~# ifconfig
eth0
Link encap:Ethernet HWaddr AA-AA-AA-AA-AA-D2

Xen2
RX packets:16 errors:0 dropped:0 overruns:0 frame:0
TX packets:8 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:1000
RX bytes:552 (552.0 b) TX bytes:816 (816.0 b)

Xen3
RX packets:12 errors:0 dropped:0 overruns:0 frame:0
TX packets:20 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:1000
RX bytes:552 (552.0 b) TX bytes:816 (816.0 b)

Xen4
individual files in /usr/share/doc/*/*copyright.

Xen5
RX packets:7 errors:0 dropped:0 overruns:0 frame:0
TX packets:12 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:1000
RX bytes:552 (552.0 b) TX bytes:816 (816.0 b)

Xen6
RX packets:8 errors:0 dropped:0 overruns:0 frame:0
TX packets:12 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:1000
RX bytes:552 (552.0 b) TX bytes:816 (816.0 b)

(none):~# ping 10.0.3.1
PING 10.0.3.1 (10.0.3.1) 56(84) bytes of data:
64 bytes from 10.0.3.1: icmp_seq=1 ttl=64 time=3.05 ms
64 bytes from 10.0.3.1: icmp_seq=2 ttl=64 time=0.230 ms
64 bytes from 10.0.3.1: icmp_seq=3 ttl=64 time=0.230 ms
--- 10.0.3.1 ping statistics ---
3 packets transmitted, 3 received, 0% packet loss, time 2003ms
rtt min/avg/max/mdev = 0.230/1.171/3.054/1.331 ms
(none):~#
```

Results

CPU consumption at boot up

field	user	system	nice	idle	wa	hw int	sw int	steal time
VNUML	10.78	20.75	0.00	60.29	0.13	0.05	0.03	0.00
NETKIT	16.03	31.60	0.00	51.84	0.34	0.04	0.15	0.00
QEMU	63.10	23.95	0.00	10.91	0.18	0.06	0.22	0.00
XEN	2.04	4.17	0.00	58.86	22.01	0.03	0.13	6.76
VMWARE	13.29	68.72	0.00	14.34	3.23	0.11	0.14	0.00
VIRTUALBOX	3.15	4.64	65.48	19.37	6.94	0.08	0.12	0.00

Memory consumption at boot up

	Consumed	Buffers	Cache	Used (B)	Used(MB)
VNUML	613,652.98	89,257.40	247,900.77	276,494.81	270.01
NETKIT	466,763.82	34,243.32	242,244.45	190,276.05	185.82
QEMU	717,314.60	18,740.75	451,260.44	247,313.41	241.52
XEN	345,967.69	11,539.29	124,614.41	209,814.00	204.90
VMWARE	643,106.22	8,102.36	341,359.51	293,644.34	286.76
VIRTUALBOX	810,683.96	2,610.21	123,162.98	684,910.77	668.86



An emulation of VoD services using virtual network environments

**Walter Fuertes and
Jorge E. López de Vergara**

**Departamento de Ingeniería Informática, Escuela Politécnica Superior
Universidad Autónoma de Madrid, Spain
email: {[walter.fuertes](mailto:walter.fuertes@uam.es), [jorge.lopez_vergara](mailto:jorge.lopez_vergara@uam.es)}@uam.es**

Motivation and Goals

□ Goals

- This paper proposes a **method to improve** the results obtained in virtual network environments, trying to resemble those obtained in real environments;
- We have emulated **video-on-demand (VoD) over ADSL** by means of the Xen;
- We have done the comparisons between virtual and real environments; then **we adjusted** the network environment parameters to **emulate the operational aspects** of the real environment.



An emulation of VoD services using virtual network environments Walter Fuertes¹ and Jorge E. López de Vergara²

¹ walter.fuertes@uam.es, <http://www.uam.es/>

² jorge.lopez.vergara@uam.es

Departamento de Ingeniería Informática, Escuela Politécnica Superior
Universidad Autónoma de Madrid, Madrid, Spain

Abstract: Virtualization platforms are a viable alternative for the implementation of IP network experimentation environments. These platforms facilitate the conducting of tests as if a real environment were used and therefore can reduce the risk of failure as well as investment and experimentation costs. This paper proposes to develop a method to improve the results obtained in virtual network environments, trying to resemble those obtained in a real environment. To carry this out, we have emulated a video-on-demand service over ADSL using Xen as a virtualization tool, just as it would have been through a real ADSL connection. Connectivity, IP addressing, switching, routing and video streaming were tested to check the functionality of virtual network environments. Then, the bandwidth, the delay, and the inter-arrival time of video streaming packets were measured both in real and virtual environments. Finally, these parameters were tuned in the virtual network environments obtaining a similar behavior in clients and servers of both cases.

Keywords: Emulation, performance measurement, traffic monitoring, video streaming, virtualization.

1 Introduction

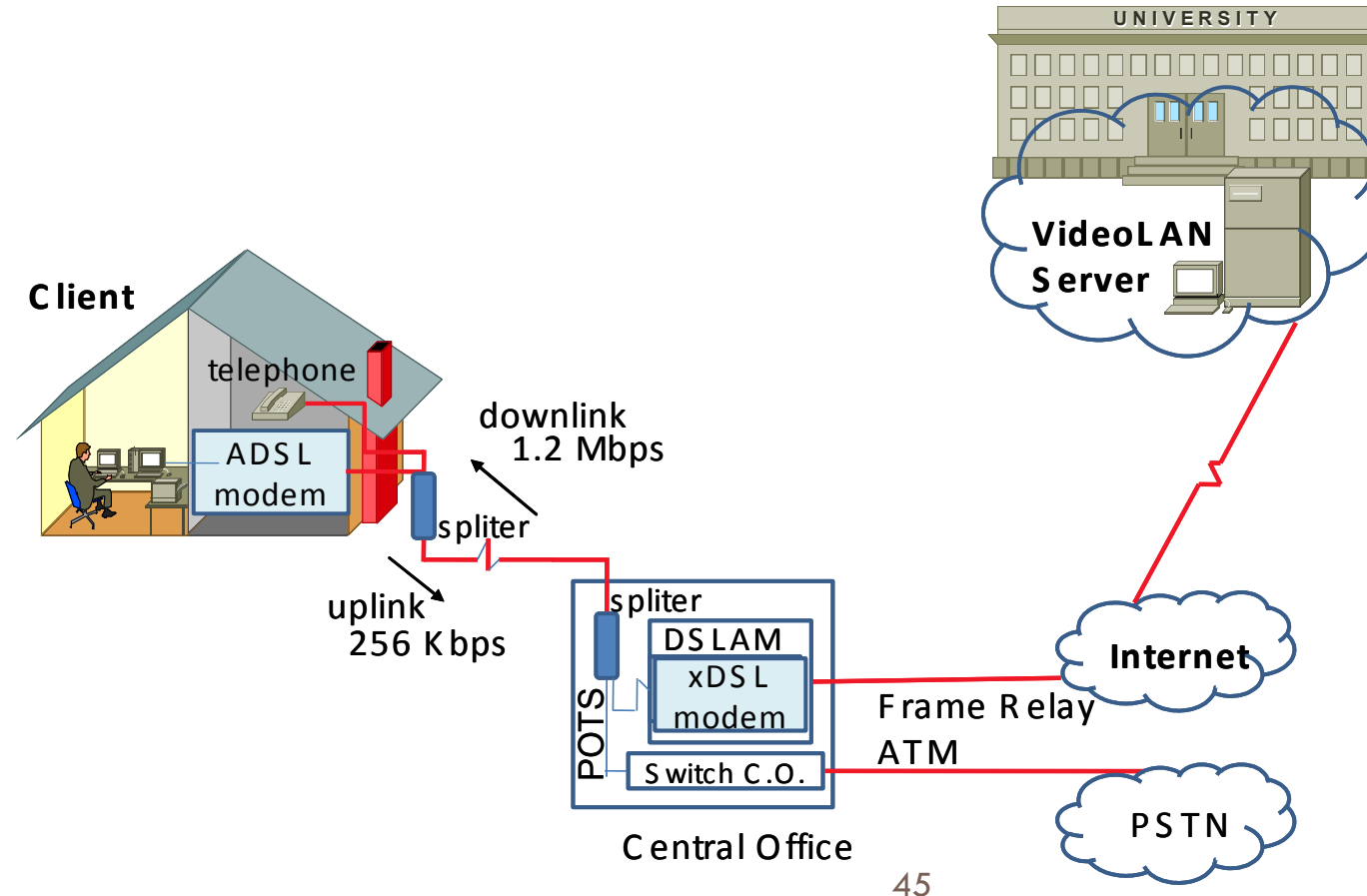
Virtualization platforms are a potential technology to reproduce a real network topology using a virtual environment. They enable interconnected equipment to be emulated, which only requires the resources of a single physical computer. Virtualization can be used to maintain, execute and test multiple software environments [MHH⁺07], and to provide facilities for network dimensioning. In addition, virtualization platforms allow the evaluation of network service provisioning prior to production; therefore they can reduce the risk of network failures, as well as investment and experimental costs.

Currently there are at least two ways to achieve the networks dimensioning and comparative performance measurements. One alternative is to set up a mock-up network infrastructure in parallel, but this would require new equipment and hardware devices making it an expensive solution. The other alternative is to use simulation tools, such as NS2 [NS2], among others, which are used to evaluate the performance of networks. Nevertheless, these simulators are trying to replicate the performance of a real system (end-to-end delay, packet loss, etc.), using software, but are not able to closely reproduce the features and behavior of the hardware in a real system (emulation of new devices, configurations, architectures) [Rim07].

Facing these two alternatives, this work proposes to use virtualization platforms to assess the network service provision. According to the results obtained in [Mun08], virtualization tech-

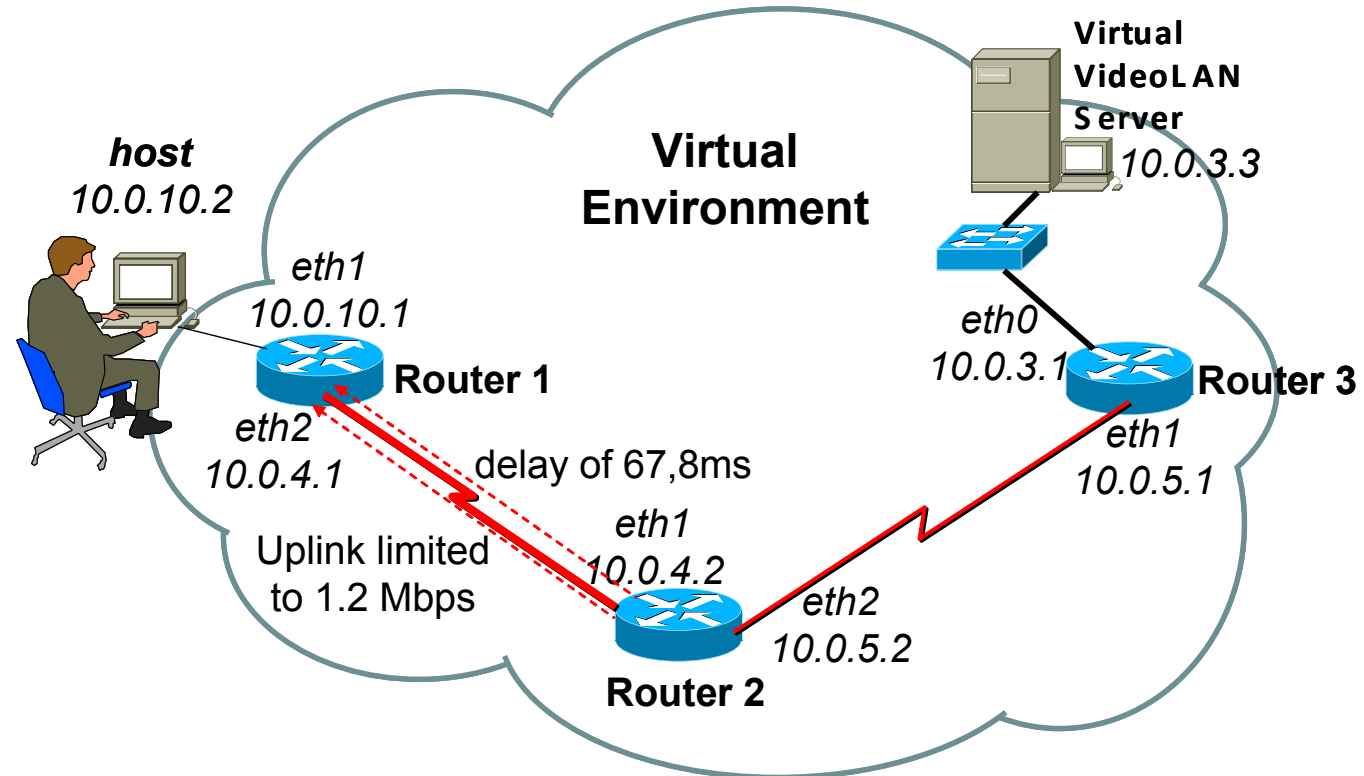
Emulation experiments

□ Step 1: Description of real environment



Emulation experiments

- **Step 2: Design and implementation of the virtual network environment**



Experimental results

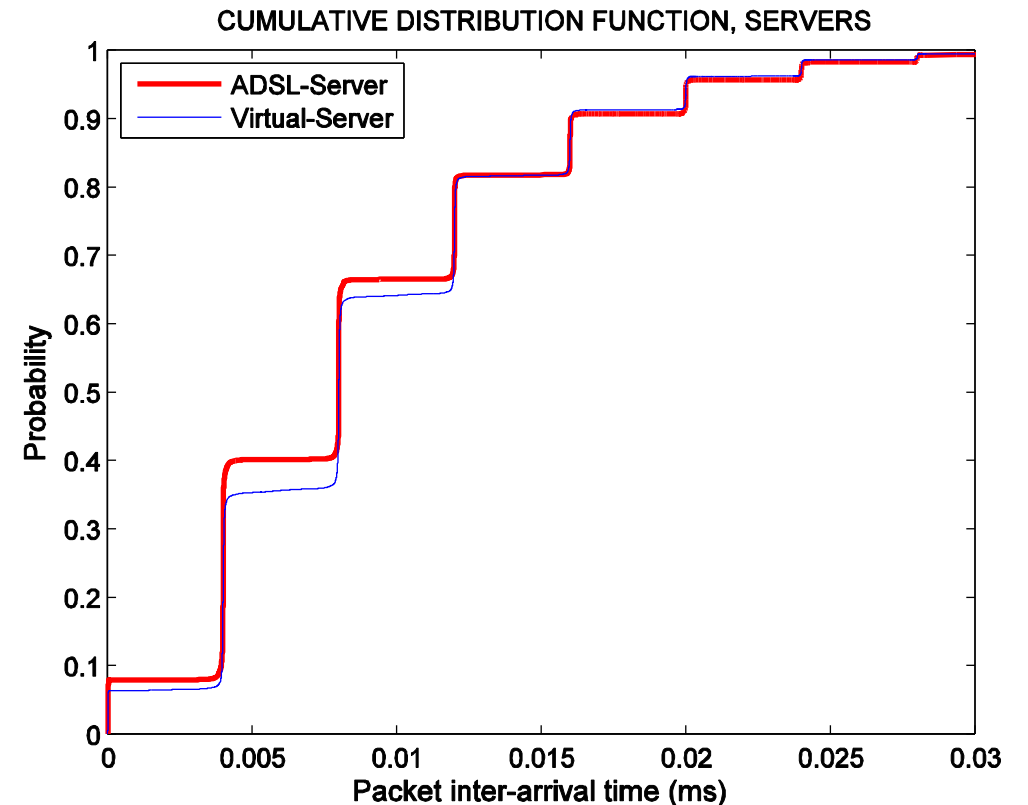
□ Comparison between the ADSL server and virtual server

Kullback-Leibler divergence

$$Divergence(R \parallel V) = \int_{-\infty}^{\infty} r(x) \log \frac{r(x)}{v(x)} dx$$

Where $r(x)$ and $v(x)$ denote the R (real ADSL client) and V (virtual client) packet interarrival random variables.

- The **divergence** for the real ADSL client and the virtual client with adjustments = **0.0473**;



TALLER N° 2. Análisis de la Literatura

48

□ Tema:

- Taller N° 2: **Análisis de la Literatura**

□ Objetivos:

- Estructurar la **línea base** del Marco Teórico de su proyecto de titulación ensamblando los antecedentes de la investigación y las bases teóricas.
- Estructurar la **línea base** del Estado del Arte de su proyecto de titulación utilizando el SLR propuesto por Bárbara Kitchenham.

□ Se pide:

- **Buscar las fuentes bibliográficas** que permitan detectar, extraer y recopilar la información de interés para construir el marco teórico pertinente a su problema de investigación.
- **Buscar las fuentes bibliográficas mediante la selección de** trabajos relacionados en el tema de su investigación que le permitan tanto sentar las bases del estado del arte como obtener información sobre las tendencias actuales y los desafíos futuros.

□ Entregable:

- Trabajo individual. Elaborar un documento en formato PDF, que contenga la línea base del Marco teórico y el Estado del Arte de su proyecto de titulación.

□ Plazo:

- Sábado 10 de octubre de 2020, 16:00 y subirlo a la plataforma virtual.



Referencias Bibliográficas

- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering—a systematic literature review. *Information and software technology*, 51(1), 7-15.
- Tapia, F.; Mora, M.Á.; Fuertes, W.; Lascano, J.E.; Toulkeridis, T., “**A Container Orchestration Development that Optimizes the Etherpad Collaborative Editing Tool through a Novel Management System**”. *Electronics* 2020, 9, 828. <https://doi.org/10.3390/electronics9050828>
- Alyssa Cadena, Franklin Gualoto, Walter Fuertes, Luis Tello-Oquendo, Roberto Andrade, Freddy Tapia, and Jenny Torres, “**Metrics and Indicators of Information Security Incident Management: A Systematic Mapping Study**”. In: Rocha Á., Pereira R. (eds) *Developments and Advances in Defense and Security. Smart Innovation, Systems and Technologies*, vol 152. Springer, Singapore. DOI https://doi.org/10.1007/978-981-13-9155-2_5
- Benavides E., Fuertes W., Sanchez S., and Sanchez M. (2020). “**Classification of Phishing Attack Solutions by Employing Deep Learning Techniques: A Systematic Literature Review**”. In: Rocha Á., Pereira R. (eds) *Developments and Advances in Defense and Security. Smart Innovation, Systems and Technologies*, vol 152. Springer, Singapore. DOI https://doi.org/10.1007/978-981-13-9155-2_5
- Roberto Hernández Sampieri et al.(2014), Carlos Fernández Collado, Pilar Baptista Lucio, *Metodología de la Investigación Científica*, Mc Graw Hill, México, ISBN: 978-1-4562-2396-0.
- Héctor Luis Ávila Baray (2006) *Introducción a la Metodología de la Investigación*, Cd. Cuauhtémoc, Chihuahua, México.
- Francisco Bijarro Hernández, *Desarrollo Estratégico para la Investigación Científica*, Universidad Autónoma de Tamaulipas, ISBN-13: 978-84-690-8111-2, N° REGISTRO: 07/76456.
- Jan Feyen , ¿Cómo elaborar propuestas de investigación?, ESPE-2010.
- Robert Day, *Cómo escribir y publicar trabajos científicos*, Tercera edición en español, Washington DC, ISBN 92 75 31598 1
- Fuertes, Walter M., and Jorge E. López de Vergara. "A quantitative comparison of virtual network environments based on performance measurements." *Proceedings of the 14th HP Software University Association Workshop, Garching, Munich, Germany*. 2007.
- Fuertes, Walter. "An emulation of VoD services using virtual network environments." *Electronic Communications of the EASST* 17 (2009).